Geophysical Journal International

Geophys. J. Int. (2015) **200,** 1449–1465 doi: 10.1093/gji/ggu476
GJI Geomagnetism, rock magnetism and palaeomagnetism

General formalism for the efficient calculation of the Hessian matrix of EM data misfit and Hessian-vector products based upon adjoint sources approach

Oleg Pankratov[1,2] and Alexey Kuvshinov[1]

[1] *Institute of Geophysics, ETH,* CH-8092 *Zurich, Switzerland. E-mail:* *oleg.pankratov@gmail.com*
[2] *Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radiowave Propagation, Russian Academy of Sciences, Troitsk, Moscow* 142190, *Russia*

Accepted 2014 December 10. Received 2014 December 8; in original form 2014 March 21

SUMMARY

3-D electromagnetic (EM) studies of the Earth have advanced significantly over the past decade. Despite a certain success of the 3-D EM inversions of real data sets, the quantitative assessment of the recovered models is still a challenging problem. It is known that one can gain valuable information about model uncertainties from the analysis of Hessian matrix. However, even with modern computational capabilities the calculation of the Hessian matrix based on numerical differentiation is extremely time consuming. Much more efficient way to compute the Hessian matrix is provided by an 'adjoint sources' methodology. The computation of Hessian matrix (and Hessian-vector products) using adjoint formulation is now well-established approach, especially in seismic inverse modelling. As for EM inverse modelling we did not find in the literature a description of the approach, which would allow EM researchers to apply this methodology in a straightforward manner to their scenario of interest. In the paper, we present formalism for the efficient calculation of the Hessian matrix using adjoint sources approach. We also show how this technique can be implemented to calculate multiple Hessian-vector products very efficiently. The formalism is general in the sense that it allows to work with responses that arise in EM problem set-ups either with natural- or controlled-source excitations. The formalism allows for various types of parametrization of the 3-D conductivity distribution. Using this methodology one can readily obtain appropriate formulae for the specific sounding methods. To illustrate the concept we provide such formulae for two EM techniques: magnetotellurics and controlled-source sounding with vertical magnetic dipole as a source.

Key words: Numerical solutions; Inverse theory; Electromagnetic theory; Geomagnetic induction.

1 INTRODUCTION

A number of rigorous 3-D frequency-domain electromagnetic (EM) inverse solutions have been developed in the past two decades both in Cartesian geometries (Mackie & Madden 1993; Newman & Alumbaugh 2000; Haber *et al.* 2004; Siripunvaraporn *et al.* 2005; Sasaki & Meju 2006; Avdeev & Avdeeva 2009; Egbert & Kelbert 2012; Zhang *et al.* 2012, among others) and spherical geometries (Koyama 2001; Kelbert *et al.* 2008; Kuvshinov & Semenov 2012). Despite a certain success of the 3-D EM analysis of frequency-domain data of different types and origin there are a number of open issues to be addressed. One of the most challenging problem is a quantification of the model uncertainty.

It is known that one can gain ample information about the inverse solution from the quadratic approximation of the penalty functional in a vicinity of the minimum. To illustrate the concept let us assume for the moment that vector $\mathbf{m} = (\sigma_1, \sigma_2, \ldots, \sigma_{N_\mathcal{M}})^T$ defines the model parametrization, where superscript T stands for transpose and $\sigma_1, \sigma_2, \ldots, \sigma_{N_\mathcal{M}}$ are conductivities in $N_\mathcal{M}$ cells of the volume where we aim to recover the conductivity distribution. By formulating the inverse problem (conductivity recovery) as a minimization problem with a penalty functional, β (which is a sum of the data misfit, β_d, and the regularization term, $\lambda\beta_r$), one can state that the local behaviour of $\beta(\mathbf{m})$ in the vicinity of a stationary point $\mathbf{m}_0$ is determined by the second-order terms of the Taylor series:

$$\beta(\mathbf{m}) \approx \beta(\mathbf{m}_0) + \frac{1}{2}\Delta\mathbf{m}^T\,\mathrm{Hess}(\mathbf{m}_0)\,\Delta\mathbf{m}, \quad \Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0, \tag{1}$$

where Hess is the (Hessian) matrix, which elements are $\frac{\partial^2 \beta}{\partial \sigma_k \partial \sigma_l}$, $k, l = 1, 2, \ldots, N_{\mathcal{M}}$, at $\mathbf{m} = \mathbf{m}_0$. Analysis of eq. (1) suggests a way to specify the bounds within which the optimal model $\mathbf{m}_0$ can be perturbed without increasing $\beta(\mathbf{m}_0)$ beyond a predefined value $\beta(\mathbf{m}_0) + \Delta\beta$, where $\Delta\beta$ is somehow relates to the noise in the data and to an uncertainty associated with the model inadequacy. More exactly, for perturbation $\Delta\mathbf{m}$ we require that the following inequality

$$\beta\left(\mathbf{m}_0 \pm \Delta\mathbf{m}\right) \leq \beta\left(\mathbf{m}_0\right) + \Delta\beta, \tag{2}$$

holds. In the case when the Hessian matrix is diagonal, the formula for the perturbation $\Delta\mathbf{m}$ reads (see Appendix D)

$$\Delta m_{k,k} \equiv \Delta\sigma_{k,k} = \sqrt{\frac{2\Delta\beta}{\mathrm{Hess}(\mathbf{m}_0)_{k,k}}}, \quad k = 1, 2, \ldots, N_{\mathcal{M}}. \tag{3}$$

It is seen from eq. (3) that the smaller $\mathrm{Hess}(\mathbf{m}_0)_{kk}$ the larger admissible perturbations of m_k, meaning that m_k is poorly constrained (resolved). Note, however, that usually non-diagonal elements of $\mathrm{Hess}(\mathbf{m}_0)$ are not zero. This complicates resolution analysis, and now, instead of eq. (3), a more sophisticated formula holds which is presented in Appendix D.

However, even with modern computational capabilities the calculation of the Hessian matrix of the misfit, Hess_{β_d}, based on numerical differentiation is forbidden due to the tremendous computational loads (note, that the evaluation of the Hessian matrix of the regularization term is usually straightforward and fast). A much more efficient way to calculate Hess_{β_d} is provided by adjoint sources approach. The computation of Hessian matrix (and Hessian-vector products) using adjoint formulation is now rather well-established approach, especially in seismic inverse modelling (Santosa & Symes 1988; Epanomeritakis *et al.* 2008; Fichtner & Trampert 2011; Metivier *et al.* 2013, among others). As for EM inverse modelling we did not find in the literature a description of the approach, which would allow EM researchers to apply this methodology in a straightforward manner to their scenario of interest. In fact, we found only one EM publication (Newman & Hoversten 2000), where computation of Hess_{β_d} using adjoint sources approach is discussed. Note, however, that Newman & Hoversten (2000) confined the discussion to the case when the data are controlled-source electric or magnetic fields.

In this paper, we present a general formalism for calculating second derivatives of the response functions and misfit. We also show how this technique can be implemented to calculate multiple Hessian-vector products very efficiently. The latter can be used in truncated Newton optimization methods (Nash 2000) for 3-D EM inversions, and more important, it opens an avenue for implementing a low-rank approximation of the Hessian matrix as it was discussed in section 2.4 in the paper by Martin *et al.* (2012). Such approximation could make a stochastic approach to 3-D EM inversion feasible. In addition, the approximation can be applied to quantify the uncertainty of a model that has been acquired by some inversion technique.

This paper is written in a manner that follows the line of presentation adopted in our previous paper (Pankratov & Kuvshinov 2010) which discusses efficient calculation of the gradient of the misfit, and which is also based on an adjoint sources approach. The paper is organized as follows. Section 2 introduces Maxwell's operators of the 3-D EM forward problem, gives definitions of polarizations, parametrizations, observation sites, and response functions and discusses the inverse problem set-up. Section 3 describes a formalism for fast calculation of the first derivatives of electric and magnetic fields, the response functions and the misfit. In many aspects the content of this section is similar to that presented in Pankratov & Kuvshinov (2010), but in this work we use a rather different way to demonstrate the results. Section 4 is a key section of the paper in which a methodology has been developed to calculate the second derivatives of the fields, responses, the Hessian of the misfit, as well as a Hessian-vector products efficiently. Finally, Section 5 provides actual formulae for two methods that are based on controlled and natural sources. In order to navigate the reader throughout the text the milestone formulae that are relevant for numerical implementation are framed. Conclusions and discussion are summarized in Section 6. The paper also includes a number of appendices which detail some results presented in the main body of the paper.

2 DEFINITIONS

2.1 Green's operators

Let us define an operator $\mathbf{G}^{\cdot\cdot}$ in the whole 3-D Euclidean (physical) space $\mathbb{R}^3$

$$\begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \mathbf{G}^{\cdot\cdot} \begin{pmatrix} \mathbf{j}^{\mathrm{imp}} \\ \mathbf{h}^{\mathrm{imp}} \end{pmatrix} \Leftrightarrow \begin{cases} \nabla \times \mathbf{H} = \sigma\mathbf{E} + \mathbf{j}^{\mathrm{imp}}, \\ \nabla \times \mathbf{E} = \mathrm{i}\omega\mu\mathbf{H} + \mathbf{h}^{\mathrm{imp}}, \\ \mathbf{E}(\mathbf{r}), \mathbf{H}(\mathbf{r}) \longrightarrow 0 \quad \text{as} \quad |\mathbf{r}| \longrightarrow \infty, \end{cases} \tag{4}$$

where $\mathbf{E}$ and $\mathbf{H}$ are electric and magnetic fields, $\mathbf{j}^{\mathrm{imp}}$ and $\mathbf{h}^{\mathrm{imp}}$ are impressed (extraneous) electric and magnetic sources, respectively, $\mathbf{r} \in \mathbb{R}^3$ is a position vector, $i = \sqrt{-1}$, $\omega = 2\pi/\mathrm{Period}$ is an angular frequency, $\sigma(\mathbf{r})$ and $\mu(\mathbf{r})$ are electric conductivity and magnetic permeability distributions in an earth's model, respectively. In this paper, we assume that $\sigma(\mathbf{r})$ is a real-valued function. All fields, $\mathbf{E}$, $\mathbf{H}$, $\mathbf{j}^{\mathrm{imp}}$ and $\mathbf{h}^{\mathrm{imp}}$, are complex-valued functions of ω and $\mathbf{r}$. In addition the fields $\mathbf{E}$ and $\mathbf{H}$ depend on σ and μ. In this paper, we study the derivatives with respect to σ only. Green's operator $\mathbf{G}^{\cdot\cdot}$ depends on functional arguments $\mathbf{j}^{\mathrm{imp}}$ and $\mathbf{h}^{\mathrm{imp}}$. Hereinafter the dependence of Green's operator on σ, $\mathbf{r}$, and ω is omitted but implied. Time dependence of fields is accounted for by $\mathrm{e}^{-\mathrm{i}\omega t}$, which reads, for example, for electric field as

$\check{\mathbf{E}}(\mathbf{r}, t) = \int \mathbf{E}(\mathbf{r}, \omega)\, \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}\omega$. At this stage we do not specify the coordinate system in $\mathbb{R}^3$; this means that $\mathbf{r}$ can be, for example, a triplet of Cartesian coordinates, (x, y, z), or a triplet of spherical coordinates, (r, θ, ϕ). As far as the column in the left-hand side (LHS) of eq. (4) contains two fields, $\mathbf{E}$ and $\mathbf{H}$, operator $\mathbf{G}^{\cdot\cdot}$ can be represented via operators $\mathbf{G}^{e\cdot}$, $\mathbf{G}^{h\cdot}$ $\mathbf{G}^{ee}$, $\mathbf{G}^{eh}$, $\mathbf{G}^{he}$ and $\mathbf{G}^{hh}$ as follows:

$$\mathbf{G}^{\cdot\cdot} = \begin{pmatrix} \mathbf{G}^{e\cdot} \\ \mathbf{G}^{h\cdot} \end{pmatrix} = \begin{pmatrix} \mathbf{G}^{ee} & \mathbf{G}^{eh} \\ \mathbf{G}^{he} & \mathbf{G}^{hh} \end{pmatrix}, \quad \mathbf{G}^{e\cdot} = \left(\mathbf{G}^{ee}, \mathbf{G}^{eh}\right), \quad \mathbf{G}^{h\cdot} = \left(\mathbf{G}^{he}, \mathbf{G}^{hh}\right), \tag{5}$$

where operators $\mathbf{G}^{e\cdot}$ and $\mathbf{G}^{h\cdot}$ are electric and magnetic components of $\mathbf{G}^{\cdot\cdot}$, operator $\mathbf{G}^{ee}$ is a restriction of $\mathbf{G}^{e\cdot}$ to electric sources, etc.

Let us introduce an EM field, $\mathbf{u}$, as

$$\mathbf{u}(\mathbf{r}, \omega) = \begin{pmatrix} \mathbf{E}(\mathbf{r}, \omega) \\ \mathbf{H}(\mathbf{r}, \omega) \end{pmatrix}, \tag{6}$$

which is a complex-valued 6-D vector. Let us denote the space of such vectors as $\mathcal{U} \cong \mathbb{C}^6$. Note that once we have chosen coordinates in 3-D space $\mathbb{R}^3$ with the following basis:

$$\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \tag{7}$$

then we naturally and unambiguously have a coordinate system and basis $\mathbf{e}'_1, \ldots, \mathbf{e}'_6$ in 6-D complex space $\mathcal{U}$

$$\mathbf{u}(\mathbf{r}, \omega) = \sum_{\alpha=1}^{6} u_\alpha \mathbf{e}'_\alpha, \tag{8}$$

saying that $\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3$ are $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ for electric fields whereas $\mathbf{e}'_4, \mathbf{e}'_5, \mathbf{e}'_6$ are $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ for magnetic fields, respectively.

2.2 Maxwell's differential operator

Let us rewrite the system of Maxwell's equations introduced by eq. (4) in a form

$$\mathbf{u} = \mathbf{G}^{\cdot\cdot}\left(\mathbf{f}^{\mathrm{imp}}\right) \Leftrightarrow \begin{cases} \mathbf{L}(\boldsymbol{\sigma}, \mathbf{u}) = \mathbf{f}^{\mathrm{imp}}, \\ \mathbf{u}(\mathbf{r}) \to 0 \quad \text{as} \quad \mathbf{r} \to \infty, \end{cases} \qquad \mathbf{f}^{\mathrm{imp}} = \begin{pmatrix} \mathbf{j}^{\mathrm{imp}} \\ \mathbf{h}^{\mathrm{imp}} \end{pmatrix}, \tag{9}$$

where $\boldsymbol{\sigma} = \{\sigma(\mathbf{r})\}$ and $\mathbf{L}(\boldsymbol{\sigma}, \mathbf{u})$ is defined as

$$\mathbf{L}(\boldsymbol{\sigma}, \mathbf{u}) = \begin{pmatrix} -\boldsymbol{\sigma}\, \mathbf{E} + \nabla \times \mathbf{H} \\ \nabla \times \mathbf{E} - \mathrm{i}\omega\mu\, \mathbf{H} \end{pmatrix}. \tag{10}$$

Expression $\mathbf{L}(\boldsymbol{\sigma}, \mathbf{u})$ can be symbolically written as $\mathbf{L}(\boldsymbol{\sigma}, \mathbf{u}) = \mathbb{L}(\boldsymbol{\sigma})\mathbf{u}$, where differential operator $\mathbb{L}(\boldsymbol{\sigma})$ is given by

$$\mathbb{L}(\boldsymbol{\sigma}) = \begin{pmatrix} -\boldsymbol{\sigma} & \nabla\times \\ \nabla\times & -\mathrm{i}\omega\mu \end{pmatrix}. \tag{11}$$

Thus the equation introduced in (9) now reads $\mathbb{L}(\boldsymbol{\sigma})\mathbf{u} = \mathbf{f}^{\mathrm{imp}}$. From the latter equation it is seen that Green's operator $\mathbf{G}^{\cdot\cdot}$ is an inverse to Maxwell's operator $\mathbb{L}(\boldsymbol{\sigma})$. Moreover, the action of Green's operator on the impressed current can be written as

$$\mathbf{G}^{\cdot\cdot}\left(\mathbf{f}^{\mathrm{imp}}\right)(\mathbf{r}) = \int_{\mathbb{R}^3} \hat{G}(\mathbf{r}, \mathbf{r}')\, \mathbf{f}^{\mathrm{imp}}(\mathbf{r}')\, \mathrm{d}v(\mathbf{r}'). \tag{12}$$

Here $\hat{G}(\mathbf{r}, \mathbf{r}')$ is a linear operator that vanishes as $|\mathbf{r}| \longrightarrow \infty$, and $\mathrm{d}v(\mathbf{r}')$ is an elementary volume. Note that numerically Green's operator can be considered as a solution of Maxwell's equations by finite differences, finite elements or by integral equations.

2.3 Polarizations/sources

Let

$$\left\{\mathbf{f}_p^{\mathrm{imp}}\right\}_{p \in \mathcal{P}}, \quad \mathcal{P} = \{1, 2, \ldots, N_{\mathcal{P}}\}, \tag{13}$$

be a set of linearly independent distributions (in space and frequency) of the impressed sources, $\mathbf{f}_p^{\mathrm{imp}}$. For example, in magnetotelluric (MT) studies, $N_{\mathcal{P}} = 2$, and $\mathbf{f}_1^{\mathrm{imp}}$ and $\mathbf{f}_2^{\mathrm{imp}}$ correspond to the plane waves of different orientations. Each $\mathbf{f}_p^{\mathrm{imp}}$ produces electric, $\mathbf{E}_p$, and magnetic, $\mathbf{H}_p$, fields that constitute EM field $\mathbf{u}_p$ that can be written via $\mathbf{G}^{\cdot\cdot}$ operator (4) as

$$\mathbf{u}_p = \mathbf{G}^{\cdot\cdot}\left(\mathbf{f}_p^{\mathrm{imp}}\right). \tag{14}$$

In addition, we introduce vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{E}}$ as

$$\hat{\mathbf{u}} = (\mathbf{u}_1, \ldots, \mathbf{u}_{N_{\mathcal{P}}}), \qquad \hat{\mathbf{E}} = (\mathbf{E}_1, \ldots, \mathbf{E}_{N_{\mathcal{P}}}). \tag{15}$$

2.4 Inversion domain and parametrization

As far as the inversion is usually done numerically, let the inversion domain, V^{inv}, be represented as

$$V^{\text{inv}} = \bigcup_{k=1}^{N_{\mathcal{M}}} V_k, \tag{16}$$

where $\{V_k\}_{k\in\mathcal{M}}$, $\mathcal{M} = \{1, \ldots, N_{\mathcal{M}}\}$, be a set of elementary volumes V_l, and within each volume V_l the conductivity be a constant $\sigma(\mathbf{r}) = \sigma_l$. We assemble this conductivity distribution in the following vector:

$$\boldsymbol{\sigma} = \left(\sigma_1, \ldots, \sigma_{N_{\mathcal{M}}}\right)^T, \tag{17}$$

and introduce model parametrization as

$$\mathbf{m} = \left(m_1, \ldots, m_{N_{\mathcal{M}}}\right)^T, \quad m_l = \nu^{-1}(\sigma_l), \quad l \in \mathcal{M}, \tag{18}$$

where function $\mathbf{m} = \boldsymbol{\nu}^{-1}(\boldsymbol{\sigma})$ can be implemented, for example, to preserve conductivity to be positive. Note that popular choice is $\mathbf{m} = \ln\boldsymbol{\sigma}$. We also remark that some volumes V_l might be cells (or combinations of cells) of 3-D part of the model. In addition, along with model vector $\mathbf{m}$ we introduce two arbitrary variations of the model vector

$$\delta\mathbf{m} = \left(\delta m_1, \ldots, \delta m_{N_{\mathcal{M}}}\right)^T, \qquad \delta\mathbf{n} = \left(\delta n_1, \ldots, \delta n_{N_{\mathcal{M}}}\right)^T. \tag{19}$$

Note that the variations of conductivities, $\delta\boldsymbol{\sigma}$ and $\delta\boldsymbol{\eta}$, are connected with the variations of model parameters, $\delta\mathbf{m}$ and $\delta\mathbf{n}$, as

$$\delta\boldsymbol{\sigma} = \boldsymbol{\nu}'(\mathbf{m})\,\delta\mathbf{m}, \qquad \delta\boldsymbol{\eta} = \boldsymbol{\nu}'(\mathbf{m})\,\delta\mathbf{n}. \tag{20}$$

Final remark of the section is that conductivity distribution in the form of eq. (17) can be written in alternative form

$$\boldsymbol{\sigma} = \sum_{l=1}^{N_{\mathcal{M}}} \sigma_l \mathbf{1}_{V_l}(\mathbf{r}), \tag{21}$$

where $\mathbf{1}_{V_l}(\mathbf{r})$ is an indicator function given by

$$\mathbf{1}_{V_l}(\mathbf{r}) = \begin{cases} 1, & \mathbf{r} \in V_l, \\ 0, & \mathbf{r} \notin V_l. \end{cases} \tag{22}$$

Similar representations can be also written for $\mathbf{m}$, $\delta\boldsymbol{\sigma}$, $\delta\boldsymbol{\eta}$, $\delta\mathbf{m}$ and $\delta\mathbf{n}$.

2.5 Observation sites, frequencies and response functions

Let

$$\Phi_g, \qquad g \in \mathcal{G} = \{1, 2, \ldots, N_{\mathcal{G}}\}, \tag{23}$$

be the experimental responses, and $N_{\mathcal{G}}$ is the number of all (at all available frequencies and sites) responses. Let $\mathbf{r}_g$, and ω_g be the spatial location and the frequency, respectively, at which the response Φ_g has been obtained.

Let $\mathcal{S}$ be a set of observation sites

$$\mathcal{S} = \left\{\mathbf{r}_g \mid g \in \mathcal{G}\right\} = \left\{\mathbf{s}_1, \ldots, \mathbf{s}_{N_{\mathcal{S}}}\right\}, \tag{24}$$

where $\mathbf{s}_1, \ldots, \mathbf{s}_{N_{\mathcal{S}}}$ are different observation sites, and $N_{\mathcal{S}}$ is the number of sites.

Let Ω be a set observation frequencies

$$\Omega = \left\{\omega_g \mid g \in \mathcal{G}\right\} = \left\{f_1, \ldots, f_{N_\Omega}\right\}, \tag{25}$$

where $f_1, \ldots, f_{N_\Omega}$ are different observation frequencies, and N_Ω is the number of frequencies. The definitions (23)–(25) are introduced in this specific way intentionally in order to stress the fact that in practice an actual set of experimental responses to be used for inversion varies with frequency and site.

As the generalization for intersite and non-holomorphic responses is possible, in this paper—for clarity of the exposition—we will consider only single-site and complex-differentiable (holomorphic) responses θ_g. For each $g \in \mathcal{G}$, the predicted response, θ_g, can be written in the following form

$$\theta_g(\mathbf{m}) = \Psi_g\left[\mathbf{u}_1(\mathbf{m}, \mathbf{r}_g, \omega_g), \mathbf{u}_2(\mathbf{m}, \mathbf{r}_g, \omega_g), \ldots, \mathbf{u}_{N_{\mathcal{P}}}(\mathbf{m}, \mathbf{r}_g, \omega_g)\right]. \tag{26}$$

2.6 Inverse problem formulation

We formulate the inverse problem of conductivity recovery as an optimization problem such that

$$\beta\,(\mathbf{m}, \lambda) \underset{\mathbf{m}}{\rightarrow} \min, \tag{27}$$

with a penalty functional

$$\beta\left(\mathbf{m},\lambda\right)=\beta_{d}\left(\mathbf{m}\right)+\lambda\beta_{r}\left(\mathbf{m}\right),\tag{28}$$

where $\lambda\beta_r(\mathbf{m})$ is a regularization term, and $\beta_d(\mathbf{m})$ is the data misfit which in general form is given by

$$\beta_d(\mathbf{m})=(\boldsymbol{\Theta}-\boldsymbol{\Phi})^{+}\mathcal{B}(\boldsymbol{\Theta}-\boldsymbol{\Phi}),\tag{29}$$

where $\boldsymbol{\Theta}(\mathbf{m})=[\theta_1(\mathbf{m}),\ldots,\theta_{N_{\mathcal{G}}}(\mathbf{m})]^T$ is the vector of the predicted responses for trial model $\mathbf{m}$, $\boldsymbol{\Phi}=(\Phi_1,\ldots,\Phi_{N_{\mathcal{G}}})^T$ is the vector of the experimental responses, $^{+}$ denotes a Hermitian conjugate, $\mathcal{B}$ is the inverse of the data covariance matrix, $\mathcal{B}=[\mathrm{Cov}(\boldsymbol{\Phi})]^{-1}$, where $\mathrm{Cov}(\boldsymbol{\Phi})=\mathtt{E}[(\boldsymbol{\Phi}-\mathtt{E}(\boldsymbol{\Phi}))(\boldsymbol{\Phi}-\mathtt{E}(\boldsymbol{\Phi}))^{+}]$, and $\mathtt{E}(\cdot)$ stands for the expected value. Usually in EM studies the data covariance matrix $\mathcal{B}$ is assumed to be diagonal, and thus expression (29) degenerates to the following form:

$$\beta_d(\mathbf{m})=\sum_{g\in\mathcal{G}}\left|\frac{\theta_g(\mathbf{m})-\Phi_g}{\Delta\Phi_g}\right|^2,\tag{30}$$

where $\Delta\Phi_g$ is an uncertainty of the experimental response. In what follows, we present general formalism to efficiently calculate the Hessian matrix of the data misfit in the form of eq. (30).

3 THE FIRST DERIVATIVES

3.1 The Gateaux differential and Frechet derivative

In this section, we discuss the definitions of the first differential (Gateaux differential) and the first derivative (Frechet derivative) of a function whose argument is an infinite-dimensional vector. The Frechet derivative is a derivative that generalizes the conventional finite-dimensional gradient vector of a scalar function (or a Jacobian matrix of a vector-valued function). The Gateaux differential is a more general concept than the Frechet derivative and is more convenient object to work with.

First, the 'Gateaux differential' of a function $\mathbf{L}(\mathbf{u})$ with respect to argument $\mathbf{u}$ in direction of vector $\mathbf{w}$ is:

$$D_{\mathrm{w}}\,\mathbf{L}\left(\mathbf{u}\right)=\left.\frac{\mathrm{d}}{\mathrm{d}h}\right|_{h=0}\mathbf{L}\left(\mathbf{u}+h\mathbf{w}\right),\tag{31}$$

where h runs over complex numbers and $\frac{\mathrm{d}}{\mathrm{d}h}|_{h=0}$ is an ordinary derivative, $\frac{\mathrm{d}\varphi}{\mathrm{d}h}|_{h=0}=\lim_{h\to0}\frac{\varphi(h)-\varphi(0)}{h}$, $h\in\mathbb{C}$. One can write the Gateaux differential in an alternative form

$$D_{\mathrm{w}}\,\mathbf{L}(\mathbf{u})=\left\langle\frac{\partial\mathbf{L}(\mathbf{u})}{\partial\mathbf{u}}\,\middle|\,\mathbf{w}\right\rangle,\tag{32}$$

where $\frac{\partial\mathbf{L}(\mathbf{u})}{\partial\mathbf{u}}$ stands for a 'Frechet derivative', and $\langle\cdot\,|\,\cdot\rangle$ denotes a contraction operation explained in Appendix A. Now, as far as vector $\mathbf{w}$ is arbitrary and the operator in the right-hand side (RHS) of eq. (32) is linear with respect to vector $\mathbf{w}$, the latter is referred to as a variation $\mathbf{w}(\mathbf{r})=\delta\mathbf{u}(\mathbf{r})$, so that eq. (32) takes the form of variation of $\mathbf{L}(\mathbf{u})$ with respect to variation of vector $\mathbf{u}$

$$\delta\mathbf{L}\left(\mathbf{u},\delta\mathbf{u}\right)=\left\langle\frac{\partial\mathbf{L}\left(\mathbf{u}\right)}{\partial\mathbf{u}}\,\middle|\,\delta\mathbf{u}\right\rangle,\tag{33}$$

that is an analogue of expression $\mathrm{d}f=f'\,(x)\,\mathrm{d}x$. In a similar way we introduce a variation of $\mathbf{u}(\boldsymbol{\sigma})$ with respect to variation of $\boldsymbol{\sigma}$

$$\delta\mathbf{u}(\boldsymbol{\sigma},\delta\boldsymbol{\sigma})=\left\langle\frac{\partial\mathbf{u}(\boldsymbol{\sigma})}{\partial\boldsymbol{\sigma}}\,\middle|\,\delta\boldsymbol{\sigma}\right\rangle.\tag{34}$$

3.2 The first derivative of EM field

Using the linearity section 2.2 of Maxwell's operator $\mathbb{L}(\boldsymbol{\sigma})$, eq. (32) reads

$$\left\langle\frac{\partial\mathbb{L}(\boldsymbol{\sigma})\mathbf{u}}{\partial\mathbf{u}}\,\middle|\,\mathbf{w}\right\rangle=\left.\frac{\mathrm{d}}{\mathrm{d}h}\right|_{h=0}[\mathbb{L}(\boldsymbol{\sigma})\mathbf{u}+h\mathbb{L}(\boldsymbol{\sigma})\mathbf{w}]=\mathbb{L}(\boldsymbol{\sigma})\mathbf{w}.\tag{35}$$

Assuming the dependence $\mathbf{u}=\mathbf{u}(\boldsymbol{\sigma})$ we rewrite the first eq. in RHS of (9) as

$$\mathbb{L}(\boldsymbol{\sigma})\mathbf{u}(\boldsymbol{\sigma})=\mathbf{f}^{\mathrm{imp}}.\tag{36}$$

We then differentiate eq. (36) as a composite function of $\boldsymbol{\sigma}$ and employ the chain rule as follows:

$$\left\langle\frac{\partial\mathbb{L}(\boldsymbol{\sigma})\mathbf{u}}{\partial\mathbf{u}}\,\middle|\,\frac{\partial\mathbf{u}}{\partial\boldsymbol{\sigma}}\right\rangle+\frac{\partial\mathbb{L}(\boldsymbol{\sigma})\mathbf{u}}{\partial\boldsymbol{\sigma}}=0.\tag{37}$$

RHS in eq. (37) is zero due to the fact that impressed source $\mathbf{f}^{\mathrm{imp}}$ does not depend on $\boldsymbol{\sigma}$. Contracting eq. (37) with an arbitrary vector $\delta\boldsymbol{\sigma}$ hereinafter referred to as a conductivity variation, and using eqs (A7), (35), (9), (B3), (5) and (34), we arrive at the first variation of EM field $\mathbf{u}$

$$\delta\mathbf{u}(\boldsymbol{\sigma}, \delta\boldsymbol{\sigma}) = \mathbf{G}^{\cdot e}\,(\delta\boldsymbol{\sigma}\,\mathbf{E})\,. \tag{38}$$

Applying eq. (38) to the pth source we have

$$\delta\mathbf{u}_p = \left\langle \frac{\partial\mathbf{u}_p}{\partial\boldsymbol{\sigma}} \middle| \delta\boldsymbol{\sigma} \right\rangle = \mathbf{G}^{\cdot e}\left(\delta\boldsymbol{\sigma}\,\mathbf{E}_p\right). \tag{39}$$

Decomposing eq. (39) into electric and magnetic parts we obtain

$$\delta\mathbf{E}_p(\boldsymbol{\sigma}, \delta\boldsymbol{\sigma}) = \mathbf{G}^{ee}\left(\delta\boldsymbol{\sigma}\,\mathbf{E}_p\right), \qquad \delta\mathbf{H}_p(\boldsymbol{\sigma}, \delta\boldsymbol{\sigma}) = \mathbf{G}^{he}\left(\delta\boldsymbol{\sigma}\,\mathbf{E}_p\right). \tag{40}$$

Note that the two latter expressions correspond to eqs (32) and (33) in paper Pankratov & Kuvshinov (2010). Remembering the form of our parametrization we obtain the first derivatives of the EM field

$$\frac{\partial\mathbf{E}_p}{\partial\sigma_l} = \mathbf{G}^{ee}\left(\mathbf{1}_{V_l}\mathbf{E}_p\right), \qquad \frac{\partial\mathbf{H}_p}{\partial\sigma_l} = \mathbf{G}^{he}\left(\mathbf{1}_{V_l}\mathbf{E}_p\right). \tag{41}$$

3.3 The first derivative of a response function

In this section, we discuss the first differential of response $\theta_g(\mathbf{m})$. This differential can be written as

$$\delta\theta_g(\delta\mathbf{m}) = \left\langle \frac{\partial\theta_g}{\partial\mathbf{m}} \middle| \delta\mathbf{m} \right\rangle. \tag{42}$$

It is possible (see eq. C4) to rewrite eq. (42) in the following form:

$$\delta\theta_g(\delta\mathbf{m}) = \left\langle \delta\boldsymbol{\sigma}\,\hat{\mathbf{E}} \middle| \mathbf{G}^{e\cdot}\left(\frac{\partial\Psi_g}{\partial\hat{\mathbf{u}}}\delta_{\mathbf{r}_g}\right)\right\rangle\bigg|_{\omega_g}, \tag{43}$$

where $\delta\boldsymbol{\sigma} = \boldsymbol{\nu}'(\mathbf{m})\,\delta\mathbf{m}$ (see eq. 20), $\hat{\mathbf{u}}$ and $\hat{\mathbf{E}}$ are defined in eq. (15), and $\delta_{\mathbf{r}_g}$ is the Dirac's delta function. Remembering the form of our parametrization we obtain the first derivative of a response function as:

$$\boxed{\frac{\partial\theta_g}{\partial m_l} = \sum_{p\in\mathcal{P}} \nu_l' \int_{V_l} \mathbf{E}_p \cdot \mathbf{G}^{e\cdot}\left(\frac{\partial\Psi_g}{\partial\mathbf{u}_p}\delta_{\mathbf{r}_g}\right)\bigg|_{\omega_g} \mathrm{d}v, \qquad \nu' = \frac{\mathrm{d}\nu}{\mathrm{d}m}.} \tag{44}$$

3.4 The gradient of a misfit

From eq. (30) we obtain the differential of the data misfit with respect to the model variation as

$$\delta\beta_d(\mathbf{m}, \delta\mathbf{m}) = 2\,\mathrm{Re}\sum_{g\in\mathcal{G}} \frac{(\theta_g(\mathbf{m}) - \Phi_g)^*}{|\Delta\Phi_g|^2}\delta\theta_g(\mathbf{m}, \delta\mathbf{m}). \tag{45}$$

Now, substituting eq. (43) into the latter equation, then changing the order of the summation and integration, and remembering our parametrization we obtain an expression for the gradient of the data misfit as

$$\boxed{\frac{\partial\beta_d}{\partial m_l} = 2\nu_l'\,\mathrm{Re}\sum_{\substack{\omega\in\Omega\\ p\in\mathcal{P}}} \int_{V_l} \mathbf{E}_p(\omega)\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)\,\mathrm{d}v,} \tag{46}$$

where an 'adjoint source' $\mathbf{J}_p^M$ is given by

$$\boxed{\mathbf{J}_p^M(\omega) = \sum_{g:\omega_g=\omega} \frac{(\theta_g - \Phi_g)^*}{|\Delta\Phi_g|^2}\frac{\partial\Psi_g}{\partial\mathbf{u}_p}\delta_{\mathbf{r}_g}\bigg|_{\omega}.} \tag{47}$$

Table 1. The steps needed to calculate the gradient of the data misfit β_d.

The term	Indices range	Number of forward modellings
$\mathbf{u}_p(\omega) = \mathbf{G}^{\cdot e}\left(\mathbf{f}_p^{\mathrm{imp}}\right)$	$p \in \mathcal{P}, \omega \in \Omega$	$N_\mathcal{P} N_\Omega$
$\mathbf{E}_p(\omega)$, $\theta_g(\mathbf{m})$, $\frac{\partial \Psi_g}{\partial \mathbf{u}_p}$, $\mathbf{J}_p^M(\omega)$		0
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)$	$p \in \mathcal{P}, \omega \in \Omega$	$N_\mathcal{P} N_\Omega$
The total number of forward modellings		$2N_\mathcal{P} N_\Omega$

Table 1 summarizes the steps needed to calculate the misfit gradient. From the eq. (46) it is seen that we need $2N_\mathcal{P} N_\Omega$ forward modellings in total to calculate the data misfit gradient.

4 THE SECOND DERIVATIVES

4.1 The second Gateaux differential

In the same line as we did in Section 3.1 for the first Gateaux differential (see eq. 34), we introduce the second Gateaux differential of a function $f(\boldsymbol{\sigma})$ as the following contraction:

$$\delta^2 f(\boldsymbol{\sigma}, \delta\boldsymbol{\sigma}, \delta\boldsymbol{\eta}) = \left\langle \delta\boldsymbol{\sigma} \,\middle|\, \frac{\partial^2 f(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}^2} \,\middle|\, \delta\boldsymbol{\eta} \right\rangle. \tag{48}$$

Function f could be either vector- or scalar-valued, and either complex- or real-valued. In the discrete case $\frac{\partial^2 f(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}^2}$ is the desired Hessian matrix if f is the data misfit function.

4.2 The second derivative of EM field

Let us differentiate eq. (37) with respect to $\boldsymbol{\sigma}$

$$\left\langle\left\langle \frac{\partial^2 \mathbb{L}(\boldsymbol{\sigma})\mathbf{u}}{\partial \mathbf{u}^2} \,\middle|\, \frac{\partial \mathbf{u}}{\partial \boldsymbol{\sigma}} \right\rangle + \frac{\partial^2 \mathbb{L}(\boldsymbol{\sigma})\mathbf{u}}{\partial \mathbf{u}\partial \boldsymbol{\sigma}} \,\middle|\, \frac{\partial \mathbf{u}}{\partial \boldsymbol{\sigma}} \right\rangle + \left\langle \frac{\partial \mathbb{L}(\boldsymbol{\sigma})\mathbf{u}}{\partial \mathbf{u}} \,\middle|\, \frac{\partial^2 \mathbf{u}}{\partial \boldsymbol{\sigma}^2} \right\rangle + \left\langle \frac{\partial \mathbf{u}}{\partial \boldsymbol{\sigma}} \,\middle|\, \frac{\partial^2 \mathbb{L}(\boldsymbol{\sigma})\mathbf{u}}{\partial \mathbf{u}\partial \boldsymbol{\sigma}} \right\rangle + \frac{\partial^2 \mathbb{L}(\boldsymbol{\sigma})\mathbf{u}}{\partial \boldsymbol{\sigma}^2} = 0. \tag{49}$$

Since expression $\mathbb{L}(\boldsymbol{\sigma})\mathbf{u}$ depends linearly on $\mathbf{u}$, $\frac{\partial^2 \mathbb{L}(\boldsymbol{\sigma})\mathbf{u}}{\partial \mathbf{u}^2} = 0$. Similarly, $\mathbb{L}(\boldsymbol{\sigma})$ depends linearly on $\boldsymbol{\sigma}$, thus $\frac{\partial^2 \mathbb{L}(\boldsymbol{\sigma})\mathbf{u}}{\partial \boldsymbol{\sigma}^2} = 0$. Then using eqs (9), (33), (38) and (B7), and contracting with arbitrary variations $\delta\boldsymbol{\sigma}(\mathbf{r})$ and $\delta\boldsymbol{\eta}(\mathbf{r})$ of the conductivity distribution, we rewrite eq. (49) as

$$\delta^2 \mathbf{u}(\delta\boldsymbol{\sigma}, \delta\boldsymbol{\eta}) = \left\langle \delta\boldsymbol{\sigma} \,\middle|\, \frac{\partial^2 \mathbf{u}}{\partial \boldsymbol{\sigma}^2} \,\middle|\, \delta\boldsymbol{\eta} \right\rangle = \mathbf{G}^{\cdot e}\left(\delta\boldsymbol{\sigma}\, \mathbf{G}^{ee}\left(\delta\boldsymbol{\eta}\, \mathbf{E}\right) + \delta\boldsymbol{\eta}\, \mathbf{G}^{ee}\left(\delta\boldsymbol{\sigma}\, \mathbf{E}\right)\right). \tag{50}$$

Applying the result to the pth source we obtain the formula to calculate the second derivative of the EM field as

$$\delta^2 \mathbf{u}_p(\boldsymbol{\sigma}, \delta\boldsymbol{\sigma}, \delta\boldsymbol{\eta}) = \mathbf{G}^{\cdot e}\left(\delta\boldsymbol{\sigma}\, \mathbf{G}^{ee}\left(\delta\boldsymbol{\eta}\, \mathbf{E}_p\right) + \delta\boldsymbol{\eta}\, \mathbf{G}^{ee}\left(\delta\boldsymbol{\sigma}\, \mathbf{E}_p\right)\right). \tag{51}$$

Remembering the form of our parametrization we obtain the second derivatives of the EM field as

$$\frac{\partial^2 \mathbf{u}_p}{\partial \sigma_k \partial \sigma_l} = \mathbf{G}^{\cdot e}\left(\mathbf{1}_{V_l} \mathbf{G}^{ee}\left(\mathbf{1}_{V_k} \mathbf{E}_p\right) + \mathbf{1}_{V_k} \mathbf{G}^{ee}\left(\mathbf{1}_{V_l} \mathbf{E}_p\right)\right). \tag{52}$$

Decomposing eq. (52) into electric and magnetic parts we have

$$\begin{cases} \frac{\partial^2 \mathbf{E}_p}{\partial \sigma_k \partial \sigma_l} = \mathbf{G}^{ee}\left(\mathbf{1}_{V_l}\, \mathbf{G}^{ee}\left(\mathbf{1}_{V_k}\, \mathbf{E}_p\right) + \mathbf{1}_{V_k}\, \mathbf{G}^{ee}\left(\mathbf{1}_{V_l}\, \mathbf{E}_p\right)\right), \\ \frac{\partial^2 \mathbf{H}_p}{\partial \sigma_k \partial \sigma_l} = \mathbf{G}^{he}\left(\mathbf{1}_{V_l}\, \mathbf{G}^{ee}\left(\mathbf{1}_{V_k}\, \mathbf{E}_p\right) + \mathbf{1}_{V_k}\, \mathbf{G}^{ee}\left(\mathbf{1}_{V_l}\, \mathbf{E}_p\right)\right). \end{cases} \tag{53}$$

4.3 The second derivative of a response function

Using the chain rule, we can write the second differential of a response as

$$\delta^2 \theta_g(\mathbf{m}, \delta\mathbf{m}, \delta\mathbf{n}) = \left\langle \delta\hat{\mathbf{u}}(\boldsymbol{\nu}'\delta\mathbf{m}) \,\middle|\, \frac{\partial^2 \Psi_g}{\partial \hat{\mathbf{u}}^2} \,\middle|\, \delta\hat{\mathbf{u}}(\boldsymbol{\nu}'\delta\mathbf{n}) \right\rangle + \left\langle \frac{\partial \Psi_g}{\partial \hat{\mathbf{u}}} \,\middle|\, \delta^2\hat{\mathbf{u}}(\boldsymbol{\nu}'\delta\mathbf{m}, \boldsymbol{\nu}'\delta\mathbf{n}) + \delta\hat{\mathbf{u}}\left(\boldsymbol{\nu}''(\mathbf{m})\, \delta\mathbf{m}\, \delta\mathbf{n}\right) \right\rangle, \tag{54}$$

where $\delta^2\hat{\mathbf{u}}(\delta\boldsymbol{\sigma},\delta\boldsymbol{\eta}) = (\delta^2\mathbf{u}_1(\delta\boldsymbol{\sigma},\delta\boldsymbol{\eta}),\ldots,\delta^2\mathbf{u}_{N_\mathcal{P}}(\delta\boldsymbol{\sigma},\delta\boldsymbol{\eta}))$. Substituting the observation site $\mathbf{r}_g$ and frequency ω_g into eq. (54) and using eqs (38), (50), (C3), and evolving the polarization summation, we write the second differential of response θ_g in a form

$$\delta^2\theta_g(\delta\mathbf{m},\delta\mathbf{n}) = \sum_{p\in\mathcal{P}}\left\langle \mathbf{G}^{e\cdot}\left(\sum_{q\in\mathcal{P}}\frac{\partial^2\Psi_g}{\partial\mathbf{u}_p\partial\mathbf{u}_q}\mathbf{G}^{\cdot e}\left(\boldsymbol{\nu}'\delta\mathbf{m}\,\mathbf{E}_q\right)\delta_{\mathbf{r}_g}\right)\middle|\,\boldsymbol{\nu}'\delta\mathbf{n}\,\mathbf{E}_p\right\rangle\Bigg|_{\omega_g}$$
$$+\sum_{p\in\mathcal{P}}\left\langle \mathbf{G}^{e\cdot}\left(\frac{\partial\Psi_g}{\partial\mathbf{u}_p}\delta_{\mathbf{r}_g}\right)\middle|\,\boldsymbol{\nu}'\delta\mathbf{m}\,\mathbf{G}^{ee}\left(\boldsymbol{\nu}'\delta\mathbf{n}\,\mathbf{E}_p\right)+\boldsymbol{\nu}'\delta\mathbf{n}\,\mathbf{G}^{ee}\left(\boldsymbol{\nu}'\delta\mathbf{m}\,\mathbf{E}_p\right)+\boldsymbol{\nu}''\delta\mathbf{m}\,\delta\mathbf{n}\,\mathbf{E}_p\right\rangle\Bigg|_{\omega_g}. \tag{55}$$

Note that each term $\frac{\partial^2\Psi_g}{\partial\mathbf{u}_p\partial\mathbf{u}_q}$ is an EM method-dependent 6×6 matrix.

4.4 The Hessian of a misfit

Using the chain rule, the second differential of the data misfit can be written as follows:

$$\delta^2\beta_d(\delta\mathbf{m},\delta\mathbf{n}) = 2\,\mathrm{Re}\sum_{g\in\mathcal{G}}\frac{1}{|\Delta\Phi_g|^2}\left((\delta\theta_g(\delta\mathbf{m}))^*\,\delta\theta_g(\delta\mathbf{n}) + (\theta_g-\Phi_g)^*\,\delta^2\theta_g(\delta\mathbf{m},\delta\mathbf{n})\right). \tag{56}$$

Substituting eqs (43) and (55) into eq. (56) we have

$$\delta^2\beta_d(\delta\mathbf{m},\delta\mathbf{n}) = \mathrm{Re} < \delta\mathbf{n}\,|\,\mathbf{F}(\delta\mathbf{m}) >,\qquad \mathbf{F}(\delta\mathbf{m}) = \mathbf{F}^A(\delta\mathbf{m}) + \mathbf{F}^L(\delta\mathbf{m}), \tag{57}$$

where

$$\mathbf{F}^A(\delta\mathbf{m}) = 2\boldsymbol{\nu}'\sum_{\substack{\omega\in\Omega\\ p\in\mathcal{P}}}\mathbf{E}_p\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^B(\delta\boldsymbol{\sigma},\omega)\right)\big|_\omega,$$

$$\mathbf{F}^L(\delta\mathbf{m}) = 2\sum_{\substack{\omega\in\Omega\\ p\in\mathcal{P}}}\left\{\boldsymbol{\nu}'\mathbf{E}_p\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^\Psi(\delta\boldsymbol{\sigma},\omega)\right)+\left(\boldsymbol{\nu}'\mathbf{G}^{ee}\left(\delta\boldsymbol{\sigma}\,\mathbf{E}_p\right)+\boldsymbol{\nu}''\delta\mathbf{m}\,\mathbf{E}_p\right)\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)+\boldsymbol{\nu}'\mathbf{E}_p\cdot\mathbf{G}^{ee}\left(\delta\boldsymbol{\sigma}\,\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)\right)\right\}\Big|_\omega, \tag{58}$$

Here the adjoint source $\mathbf{J}_p^M$ is defined in eq. (47), and new adjoint sources $\mathbf{J}_p^B$, $\mathbf{J}_p^\Psi$ are as follows:

$$\mathbf{J}_p^B(\delta\boldsymbol{\sigma},\omega) = \sum_{g:\omega_g=\omega}\sum_{q\in\mathcal{P}}\frac{1}{|\Delta\Phi_g|^2}\left\langle\delta\boldsymbol{\sigma}\,\mathbf{E}_q\,\middle|\,\mathbf{G}^{e\cdot}\left(\frac{\partial\Psi_g}{\partial\mathbf{u}_q}\delta_{\mathbf{r}_g}\right)\right\rangle^*\frac{\partial\Psi_g}{\partial\mathbf{u}_p}\delta_{\mathbf{r}_g}\Bigg|_\omega, \tag{59}$$

$$\mathbf{J}_p^\Psi(\delta\boldsymbol{\sigma},\omega) = \sum_{g:\omega_g=\omega}\sum_{q\in\mathcal{P}}\frac{(\theta_g-\Phi_g)^*}{|\Delta\Phi_g|^2}\frac{\partial^2\Psi_g}{\partial\mathbf{u}_p\partial\mathbf{u}_q}\mathbf{G}^{\cdot e}\left(\delta\boldsymbol{\sigma}\,\mathbf{E}_q\right)\delta_{\mathbf{r}_g}\Bigg|_\omega. \tag{60}$$

Note that the adjoint sources, $\mathbf{J}_p^M$, $\mathbf{J}_p^B$ and $\mathbf{J}_p^\Psi$, are corresponding dipoles at the observation sites, $\mathbf{r}_g$. Remembering the form of our parametrization we obtain an expression for the Hessian matrix of the misfit

$$\frac{\partial^2\beta_d}{\partial m_k\partial m_l} = \mathrm{Re}\left(\mathcal{H}_{kl}^A+\mathcal{H}_{kl}^L\right), \tag{61}$$

where

$$\mathcal{H}_{kl}^A = \langle\mathbf{1}_{V_k}\,|\,\mathbf{F}^A(\mathbf{1}_{V_l})\rangle,\qquad \mathcal{H}_{kl}^L = \langle\mathbf{1}_{V_k}\,|\,\mathbf{F}^L(\mathbf{1}_{V_l})\rangle. \tag{62}$$

Evolving the angle-brackets contraction into 3-D integral expressions, we obtain the final formulae for $\mathcal{H}_{kl}^A$ and $\mathcal{H}_{kl}^L$

$$\mathcal{H}_{kl}^A = 2\sum_{\substack{\omega\in\Omega\\ p\in\mathcal{P}}}\int_{V_l}\nu_l'\nu_k'\mathbf{E}_p\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^B(\mathbf{1}_{V_k},\omega)\right)\mathrm{d}v\big|_\omega,$$

$$\mathcal{H}_{kl}^L = 2\sum_{\substack{\omega\in\Omega\\ p\in\mathcal{P}}}\int_{V_l}\left\{\nu_l'\nu_k'\mathbf{E}_p\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^\Psi(\mathbf{1}_{V_k},\omega)\right)+\left(\nu_l'\nu_k'\mathbf{G}^{ee}\left(\mathbf{1}_{V_k}\mathbf{E}_p\right)+\delta_{lk}\nu_k''\mathbf{E}_p\right)\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)+\nu_l'\nu_k'\mathbf{E}_p\cdot\mathbf{G}^{ee}\left(\mathbf{1}_{V_k}\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)\right)\right\}\mathrm{d}v\Big|_\omega.$$

We make here four notes.

(i) Term $\int_{V_l}\mathbf{E}_p\cdot\mathbf{G}^{e\cdot}(\mathbf{J}_p^\Psi(\mathbf{1}_{V_k}))\,\mathrm{d}v$ vanishes if the response Ψ is a linear function of EM field $\mathbf{u}$ (e.g. for most of the CSEM methods).
(ii) Term $\int_{V_l}\delta_{lk}\nu_l''\mathbf{E}_p(\omega)\cdot\mathbf{G}^{e\cdot}(\mathbf{J}_p^M(\omega))\,\mathrm{d}v$ vanishes if $\boldsymbol{\sigma}=\mathbf{m}$.

Table 2. The steps needed to calculate the Hessian of data misfit β_d.

The term	Indices range	Number of forward modellings
$\mathbf{u}_p(\omega) = \mathbf{G}^{\cdot e}\left(\mathbf{f}_p^{\mathrm{imp}}(\omega)\right)$	$p \in \mathcal{P}, \omega \in \Omega$	$N_{\mathcal{P}} N_\Omega$
$\mathbf{E}_p(\omega), \theta_g(\mathbf{m}), \frac{\partial \Psi_g}{\partial \mathbf{u}_p}$		0
$\mathbf{G}^{\cdot\cdot}\left(\mathbf{e}'_\alpha \delta_{\mathbf{s}_a}\right)\big\|_\omega$	$\alpha = 1, \ldots, 6, \mathbf{s}_a \in \mathcal{S}, \omega \in \Omega$	$\tilde{N}\ (\leq 6 N_{\mathcal{S}} N_\Omega)$
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^B(\mathbf{1}_{V_k}, \omega)\right)$,		0
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)$		0
$\mathbf{G}^{ee}\left(\mathbf{1}_{V_k}\, \mathbf{e}_\alpha\right)\big\|_\omega$	$\alpha = 1, 2, 3, k \in \mathcal{M}, \omega \in \Omega$	$3 N_{\mathcal{M}} N_\Omega$
$\mathbf{G}^{ee}\left(\mathbf{1}_{V_k}\, \mathbf{E}_p(\omega)\right)$		0
$\mathbf{G}^{ee}\left(\mathbf{1}_{V_k}\, \mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)\right)$		0
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^\Psi(\mathbf{1}_{V_k}, \omega)\right)$		0
The total number of forward modellings		$N_{\mathcal{P}} N_\Omega + \tilde{N} + 3 N_{\mathcal{M}} N_\Omega$

(iii) One can readily generalize the concept for complex-valued conductivity. The corresponding formulas are provided in Appendix E.

(iv) In the context of optimization, by neglecting the second term in RHS of eq. (61), one arrives to approximation of Hessian which is used, for example, in Gauss–Newton method.

Table 2 summarizes the steps needed to calculate the Hessian matrix. These steps are as follows:

(i) First, we make $N_{\mathcal{P}} N_\Omega$ forward runs to calculate EM fields for all the sources and frequencies.

(ii) Second, we make $\tilde{N}$ forward runs to calculate $\mathbf{G}^{\cdot\cdot}(\mathbf{e}'_\alpha \delta_{\mathbf{s}_a})$ for those α, a, ω that are involved in eqs (64)–(66). Note that for many applications $\tilde{N}$ is significantly smaller than $6 N_{\mathcal{S}} N_\Omega$, for example due to distribution of the responses over observation sites and frequencies or due to the fact that only a few of EM field components are involved in the expressions for the responses. At any case, $\tilde{N} \leq 6 N_{\mathcal{S}} N_\Omega$.

(iii) Third, we calculate $\mathbf{G}^{\cdot e}(\mathbf{1}_{V_k}\, \mathbf{e}_\alpha)$ for $\alpha = 1, 2, 3$, $k \in \mathcal{M}$, $\omega \in \Omega$;

(iv) Fourth, we calculate $\mathbf{G}^{e\cdot}(\mathbf{J}_p^M(\omega))$, $\mathbf{G}^{e\cdot}(\mathbf{J}_p^B(\mathbf{1}_{V_k}, \omega))$, $\mathbf{G}^{e\cdot}(\mathbf{J}_p^\Psi(\mathbf{1}_{V_k}, \omega))$, $\mathbf{G}^{ee}(\mathbf{1}_{V_k}\, \mathbf{G}^{e\cdot}(\mathbf{J}_p^M(\omega)))$ for $k \in \mathcal{M}$, $\omega \in \Omega$, $p \in \mathcal{P}$ for none of forward runs using the fields calculated at the previous steps and using the following formulae

$$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right) = \sum_{g:\omega_g=\omega} \frac{(\theta_g - \Phi_g)^*}{|\Delta\Phi_g|^2} \mathbf{G}^{e\cdot}\left(\frac{\partial \Psi_g}{\partial \mathbf{u}_p} \delta_{\mathbf{r}_g}\right)\Bigg|_\omega . \tag{64}$$

$$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^B(\mathbf{1}_{V_k}, \omega)\right) = \sum_{g:\omega_g=\omega} \sum_{q \in \mathcal{P}} \frac{1}{|\Delta\Phi_g|^2} \left\langle \mathbf{1}_{V_k} \mathbf{E}_q \,\middle|\, \mathbf{G}^{e\cdot}\left(\frac{\partial \Psi_g}{\partial \mathbf{u}_q} \delta_{\mathbf{r}_g}\right)\right\rangle^* \mathbf{G}^{e\cdot}\left(\frac{\partial \Psi_g}{\partial \mathbf{u}_p} \delta_{\mathbf{r}_g}\right)\Bigg|_\omega , \tag{65}$$

$$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^\Psi(\mathbf{1}_{V_k}, \omega)\right) = \sum_{g:\omega_g=\omega} \sum_{q \in \mathcal{P}} \frac{(\theta_g - \Phi_g)^*}{|\Delta\Phi_g|^2} \mathbf{G}^{e\cdot}\left(\frac{\partial^2 \Psi_g}{\partial \mathbf{u}_p \partial \mathbf{u}_q} \mathbf{G}^{\cdot e}\left(\mathbf{1}_{V_k}\, \mathbf{E}_q\right) \delta_{\mathbf{r}_g}\right)\Bigg|_\omega . \tag{66}$$

(v) Finally, we get the Hessian from eqs (59) to (63).

We see that we need totally $N_{\mathcal{P}} N_\Omega + \tilde{N} + 3 N_{\mathcal{M}} N_\Omega$ forward modellings to get the full Hessian matrix.

4.5 The Hessian-vector products

It is known that the inverse Hessian operator plays a crucial role in the parameter reconstruction. The truncated Newton methods allow one to better account for this operator (Nash 2000). These methods are based on the computation of the Newton descent direction by solving the corresponding linear system

$$\mathrm{Hess}^{(n)}\, \Delta\mathbf{m}^{(n)} = -\nabla \beta^{(n)}, \tag{67}$$

through an iterative procedure such as the conjugate gradient method. Here n is a number of iteration. The large-scale nature of 3-D EM problems requires one, however, to carefully implement this method to avoid prohibitive computational costs. In particular, this requires the capability of computing efficiently Hessian-vector products (here under *Hessian-vector product* we understand a product of the Hessian of misfit, Hess_{β_d}, and an arbitrary vector). Table 3 provides a number of forward modellings and Fig. 1 shows a flowchart to calculate this product for K times. As it is seen from the table, single Hessian-vector product can be calculated for a price of $O[N_\Omega(N_{\mathcal{P}} + N_{\mathcal{S}})]$ forward problem runs. Moreover, if such a product is calculated multiple times, as in the case of the Krylov subspace iterations, the price drops down

Table 3. The steps needed to calculate the Hessian-vector products $\mathrm{Hess}_{\beta_d}\,\mathbf{a}_k$, $k = 1, \ldots, K$.

The term	Indices range	Number of forward modellings
$\mathbf{u}_p(\omega) = \mathbf{G}^{\cdot e}\left(\mathbf{f}_p^{\mathrm{imp}}(\omega)\right)$	$p \in \mathcal{P}, \omega \in \Omega$	$N_{\mathcal{P}} N_\Omega$
$\mathbf{E}_p(\omega), \theta_g(\mathbf{m}), \frac{\partial \Psi_g}{\partial \mathbf{u}_p}$		0
$\mathbf{G}^{\cdot\cdot}\left(\mathbf{e}'_\alpha \delta_{\mathbf{s}_a}\right)\big\|_\omega$	$\alpha = 1, \ldots, 6, \mathbf{s}_a \in \mathcal{S}, \omega \in \Omega$	$\tilde{N}\ (\leq 6 N_{\mathcal{S}} N_\Omega)$
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^B(\mathbf{a}_k, \omega)\right)$		0
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)$		0
$\mathbf{G}^{\cdot e}(\mathbf{a}_k\,\mathbf{E}_p), \mathbf{G}^{ee}\left(\mathbf{a}_k\,\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M\right)\right)$	$p \in \mathcal{P}, k = 1, \ldots, K, \omega \in \Omega$	$2K N_{\mathcal{P}} N_\Omega$
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^\Psi(\mathbf{a}_k, \omega)\right)$		0
The total number of forward modellings		$N_{\mathcal{P}} N_\Omega + \tilde{N} + 2K N_{\mathcal{P}} N_\Omega$

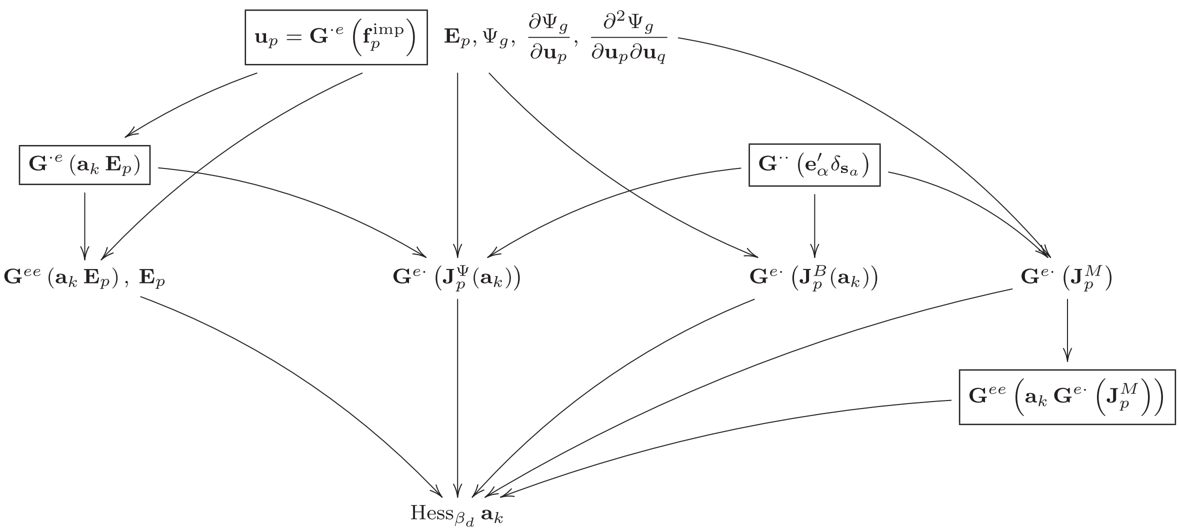

Figure 1. Flowchart of a calculation of the Hessian-vector products $\mathrm{Hess}_{\beta_d}\,\mathbf{a}_k$ for an arbitrary collection of vectors $\mathbf{a}_k$, $k = 1, \ldots, K$. The framed terms are those that require forward modellings. To calculate the terms $\mathbf{u}_p = \mathbf{G}^{\cdot e}(\mathbf{f}_p^{\mathrm{imp}})$, one needs $N_{\mathcal{P}} N_\Omega$ forward runs; these terms arise from the polarizatons $\mathbf{f}_p^{\mathrm{imp}}$ of the initial problem. To calculate the terms $\mathbf{G}^{\cdot e}(\mathbf{a}_k\,\mathbf{E}_p)$, one needs $K N_{\mathcal{P}} N_\Omega$ forward runs, and to calculate the terms $\mathbf{G}^{ee}(\mathbf{a}_k\,\mathbf{G}^{e\cdot}(\mathbf{J}_p^M))$, one needs $K N_{\mathcal{P}} N_\Omega$ forward runs; either terms arise from the vectors $\mathbf{a}_k$ for which we calculate the Hessian-vector product. To calculate the terms $\mathbf{G}^{\cdot\cdot}(\mathbf{e}'_\alpha \delta_{\mathbf{s}_a})$, one needs $\tilde{N} \leq 6 N_{\mathcal{S}} N_\Omega$ forward runs; these terms arise from the observation site locations $\mathbf{s}_a$ where one has the responses. The computational sequence is summed up in Table 3. The full Hessian matrix is obtained provided that $\mathbf{a}_k = \mathbf{1}_{V_k}$, $k \in \mathcal{M}$.

to $2N_{\mathcal{P}} N_\Omega$ per product. This is due to the fact that only the last step should be updated. Note that other application of multiple Hessian-vector products is an uncertainty quantification of a given model (the model could have been acquired by some inversion technique), using a low-rank approximation of the Hessian [see section 2.4 of Martin *et al.* (2012)].

We also notice that computation of Hessian is nothing but calculation of Hessian-vector product for $K = N_{\mathcal{M}}$ times with respective vectors $\mathbf{a}_k = \mathbf{1}_{V_k}$, $k = 1, \ldots, N_{\mathcal{M}}$.

5 EXAMPLES

5.1 A controlled-source (VMD) example

In this section, we show how the developed formalism works for the vertical magnetic dipole (VMD) sounding in the variant of a transmitter moving synchronously with a receiver. We assume that the transmitter (source) is represented by a vertical magnetic dipole of unit moment located at $\hat{\mathbf{r}}_p$

$$\mathbf{h}_p(\mathbf{r}, \omega) = \mathbf{e}_z \delta(\mathbf{r} - \hat{\mathbf{r}}_p), \quad p \in \mathcal{P} = \{1, \ldots, N_{\mathcal{P}}\}, \quad \mathbf{e}_z = \mathbf{e}_3. \tag{68}$$

We work in Cartesian coordinates x, y, z, and $\mathbf{e}_\alpha$ are unit vectors of these coordinates. The experimental response is the vertical magnetic field $H_z(\mathbf{s}_a, f_j)$ measured at all the observation sites $\mathbf{s}_a$ and frequencies f_j, $a = 1, \ldots, N_{\mathcal{S}}$, $j = 1, \ldots, N_\Omega$. As the transmitter and receiver move synchronously, the number of polarizations is equal to the number of the observation sites, $N_{\mathcal{P}} = N_{\mathcal{S}}$, and the total number of responses is

Table 4. The steps needed to calculate the Hessian-vector products $\mathrm{Hess}_{\beta_d}\,\mathbf{a}_k$, $k = 1, \ldots, K$, in the VMD case.

The term	Indices range	Number of forward modellings
$\mathbf{u}_p(\omega) = \mathbf{G}^{\cdot h}\left(\mathbf{e}_z\,\delta_{\hat{\mathbf{r}}_p}\right)\Big\|_\omega$ (transmitters)	$p = 1, \ldots, N_{\mathcal{S}},\ \omega \in \Omega$	$N_{\mathcal{S}} N_\Omega\ (N_{\mathcal{P}} = N_{\mathcal{S}})$
$\mathbf{E}_p(\omega),\ \Psi_g(\hat{\mathbf{u}}_p(\omega))$		0
$\mathbf{G}^{eh}\left(\mathbf{e}_z\,\delta_{\mathbf{s}_p}\right)\big\|_\omega$ (observation sites)	$p = 1, \ldots, N_{\mathcal{S}},\ \omega \in \Omega$	$N_{\mathcal{S}} N_\Omega\ (N_{\mathcal{P}} = N_{\mathcal{S}})$
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right),\ \mathbf{G}^{e\cdot}\left(\mathbf{J}_p^B(\mathbf{a}_k, \omega)\right)$		0
$\mathbf{G}^{\cdot\, e}(\mathbf{a}_k\,\mathbf{E}_p),\ \mathbf{G}^{ee}\left(\mathbf{a}_k\,\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M\right)\right)$	$p = 1, \ldots, N_{\mathcal{S}},\ k = 1, \ldots, K,\ \omega \in \Omega$	$2K\,N_{\mathcal{S}} N_\Omega$
The total number of forward modellings		$2(K+1) N_{\mathcal{S}} N_\Omega$

$N_{\mathcal{G}} = N_{\mathcal{S}} N_\Omega$. We enumerate the responses using index g that runs from 1 to $N_{\mathcal{G}}$ as $g(a, j) = a + N_{\mathcal{S}}(j-1)$. Next, we write the predicted responses Ψ_g and the experimental responses Φ_g as

$$\theta_g = H_z^{\mathrm{pred}}(\mathbf{s}_a, f_j), \quad \Phi_g = H_z^{\mathrm{exp}}(\mathbf{s}_a, f_j). \tag{69}$$

To calculate the adjoint sources we need all $\frac{\partial \Psi_g}{\partial \mathbf{u}_p}$ and $\frac{\partial^2 \Psi_g}{\partial \mathbf{u}_p \partial \mathbf{u}_p}$, so that using eq. (69), we have for all a, p, j

$$\frac{\partial \Psi_{g(a,j)}}{\partial \mathbf{u}_p} = \begin{cases} \mathbf{e}_6' & \text{for } p = a \\ 0 & \text{for } p \neq a \end{cases}, \qquad \mathbf{e}_6' = (0, 0, 0, 0, 0, 1)^T, \tag{70}$$

and $\frac{\partial^2 \Psi_{g(a,j)}}{\partial \mathbf{u}_p \partial \mathbf{u}_p} = 0$. Then using eqs (47), (59), (60), (61) and (63) we obtain for elements of the Hessian matrix

$$\frac{\partial^2 \beta_d}{\partial m_k \partial m_l} = \mathrm{Re}\left(\mathcal{H}_{sl}^A + \mathcal{H}_{sl}^L\right), \tag{71}$$

where

$$\mathcal{H}_{kl}^A = 2 \sum_{j=1}^{N_\Omega} \sum_{p=1}^{N_{\mathcal{P}}} \int_{V_l} \nu_k' \nu_l' \mathbf{E}_p \cdot \mathbf{G}^{e\cdot}\left(\mathbf{J}_p^B(\mathbf{1}_{V_k})\right) \mathrm{d}v \Big|_{f_j},$$

$$\mathcal{H}_{kl}^L = 2 \sum_{j=1}^{N_\Omega} \sum_{p=1}^{N_{\mathcal{P}}} \int_{V_l} \left\{\left(\nu_k' \nu_l' \mathbf{G}^{ee}\left(\mathbf{1}_{V_k}\,\mathbf{E}_p\right) + \delta_{kl} \nu_k'' \mathbf{E}_p\right) \cdot \mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M\right) + \nu_k' \nu_l' \mathbf{E}_p \cdot \mathbf{G}^{ee}\left(\mathbf{1}_{V_k}\,\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M\right)\right)\right\} \mathrm{d}v \Big|_{f_j}, \tag{72}$$

and where

$$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(f_j)\right) = \frac{(\theta_g - \Phi_g)^*}{|\Delta \Phi_g|^2} \mathbf{G}^{eh}\left(\mathbf{e}_z\,\delta_{\mathbf{s}_p}\right), \qquad g = g(p, j) \qquad p = 1, \ldots, N_{\mathcal{P}}, \qquad j = 1, \ldots, N_\Omega, \tag{73}$$

$$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^B(\mathbf{a}_k, f_j)\right) = \frac{1}{|\Delta \Phi_g|^2} \mathbf{G}^{eh}\left(\mathbf{e}_z\,\delta_{\mathbf{s}_p}\right) \left\langle \boldsymbol{\nu}' \mathbf{a}_k\,\mathbf{E}_p(f_j)\,|\,\mathbf{G}^{eh}\left(\mathbf{e}_z\,\delta_{\mathbf{s}_p}\right)\right\rangle^*, \tag{74}$$

$$\mathbf{E}_p(f_j) = \mathbf{G}^{eh}\left(\mathbf{e}_z \delta_{\hat{\mathbf{r}}_p}\right). \tag{75}$$

Table 4 provides a number of forward modellings needed to calculate Hessian-vector product K times. As it is seen from the table, single Hessian-vector product can be calculated for a price of $4N_{\mathcal{S}} N_\Omega$ forward problem runs. Moreover, if such a product is calculated multiple times the price drops down to $2N_{\mathcal{S}} N_\Omega$ per product.

5.2 An MT example

In this section, we show how the developed formalism works for the MT response functions. The number of polarizations is $N_{\mathcal{P}} = 2$, and the responses are four elements of the impedance 2×2-matrix

$$\boldsymbol{\Psi} \equiv \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} = \hat{\mathbf{E}}_\tau \hat{\mathbf{H}}_\tau^{-1} = \hat{\mathbf{E}}_\tau \mathbf{Q}, \tag{76}$$

where

$$\hat{\mathbf{E}}_\tau = [\mathbf{E}_{\tau 1}, \mathbf{E}_{\tau 2}] = \begin{pmatrix} E_{x1} & E_{x2} \\ E_{y1} & E_{y2} \end{pmatrix}, \qquad \hat{\mathbf{H}}_\tau = [\mathbf{H}_{\tau 1}, \mathbf{H}_{\tau 2}] = \begin{pmatrix} H_{x1} & H_{x2} \\ H_{y1} & H_{y2} \end{pmatrix}, \tag{77}$$

and

$$\mathbf{Q} \equiv \begin{pmatrix} Q_{1x} & Q_{1y} \\ Q_{2x} & Q_{2y} \end{pmatrix} = \hat{\mathbf{H}}_\tau^{-1} = \frac{1}{D} \operatorname{adj} \hat{\mathbf{H}}_\tau = \frac{1}{D} \begin{pmatrix} H_{y2} & -H_{x2} \\ -H_{y1} & H_{x1} \end{pmatrix}, \qquad D = \det \hat{\mathbf{H}}_\tau = \begin{vmatrix} H_{x1} & H_{x2} \\ H_{y1} & H_{y2} \end{vmatrix}. \tag{78}$$

Table 5. The steps needed calculate the Hessian-vector products $\mathrm{Hess}_{\beta_d}\,\mathbf{a}_k$, $k = 1, \ldots, K$ in the MT case.

The term	Indices range	Number of forward modellings
$\mathbf{u}_p(\omega) = \mathbf{G}^{\cdot e}\left(\mathbf{f}_p^{\mathrm{imp}}(\omega)\right)$	$p = 1, 2, \omega \in \Omega$	$2N_\Omega$
$\mathbf{E}_p(\omega)$, $Z_{\xi\eta}^{\mathrm{pred}}$, $\frac{\partial Z_{\xi\eta}^{\mathrm{pred}}}{\partial \mathbf{u}_p}$, $\frac{\partial^2 Z_{\xi\eta}^{\mathrm{pred}}}{\partial \mathbf{u}_p \partial \mathbf{u}_q}$		0
$\mathbf{G}^{\cdot\cdot}\left(\mathbf{e}'_\alpha \delta_{\mathbf{s}_a}\right)$	$\alpha = 1, 2, 4, 5, \mathbf{s}_a \in \mathcal{S}, \omega \in \Omega$	$4N_{\mathcal{S}} N_\Omega$
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^B(\mathbf{a}_k, \omega)\right)$		0
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)$		0
$\mathbf{G}^{\cdot e}(\mathbf{a}_k\,\mathbf{E}_p)$, $\mathbf{G}^{ee}\left(\mathbf{a}_k\,\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M\right)\right)$	$p = 1, 2, k = 1, \ldots, K, \omega \in \Omega$	$4KN_\Omega$
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^\Psi(\mathbf{a}_k, \omega)\right)$		0
The total number of forward modellings		$2N_\Omega + 4N_{\mathcal{S}} N_\Omega + 4K N_\Omega$

For the MT impedance the formulae for $\frac{\partial Z_{\xi\eta}}{\partial \mathbf{u}_p}$ looks as follows:

$$\frac{\partial Z_{\xi\eta}}{\partial \mathbf{u}_p} = \sum_{\alpha=1,2}\left(\delta_{\xi\alpha} Q_{q\eta}\mathbf{e}'_\alpha - Z_{\xi\alpha} Q_{q\eta}\mathbf{e}'_{\alpha+3}\right), \qquad \text{for } \xi, \eta, \in \{`x', `y'\}, \quad p \in \{1, 2\}. \tag{79}$$

Here we imply equating the indices x, y with indices 1, 2 in the subscripts. Thus the adjoint field $\mathbf{G}^{e\,\cdot}(\mathbf{J}^M)$ from eq. (64) degenerates to

$$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right) = \sum_{\substack{a\in\mathcal{S},\ \alpha=1,2\\ \xi,\eta=x,y}} \left.\frac{\left(Z_{\xi\eta}^{\mathrm{pred}} - Z_{\xi\eta}^{\mathrm{exp}}\right)^*}{|\Delta Z_{\xi\eta}^{\mathrm{exp}}|^2}\right|_{\mathbf{s}_a,\omega} \left.\left(\delta_{\xi\alpha} Q_{p\eta}\mathbf{G}^{ee}\left(\mathbf{e}_\alpha\delta_{\mathbf{s}_a}\right) - Z_{\xi\alpha} Q_{p\eta}\mathbf{G}^{eh}\left(\mathbf{e}_\alpha\delta_{\mathbf{s}_a}\right)\right)\right|_\omega, \tag{80}$$

that is equivalent to formulae (69), (70), (73) and (74) from Pankratov & Kuvshinov (2010). Next, the adjoint field $\mathbf{G}^{e\,\cdot}(\mathbf{J}^B)$ from eq. (65) is as follows:

$$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^B(\boldsymbol{\nu}'\mathbf{1}_{V_k}, \omega)\right) = \sum_{\substack{a\in\mathcal{S}\\ \xi,\eta=x,y}} \mathbf{G}^{e\cdot}\left.\left(\frac{1}{|\Delta Z_{\xi\eta}^{\mathrm{exp}}|^2}\frac{\partial Z_{\xi\eta}}{\partial \mathbf{u}_p}\delta_{\mathbf{s}_a}\right)\right|_\omega \cdot \sum_{q\in\{1,2\}} \left.\left\langle \boldsymbol{\nu}'\mathbf{1}_{V_k}\,\mathbf{E}_q \middle| \mathbf{G}^{e\cdot}\left(\frac{\partial Z_{\xi\eta}}{\partial \mathbf{u}_q}\delta_{\mathbf{s}_a}\right)\right\rangle^*\right|_\omega, \tag{81}$$

where $\frac{\partial Z_{\xi\eta}}{\partial \mathbf{u}_p}$ is from eq. (79). The second derivatives of $Z_{\xi\eta}$ are

$$\frac{\partial^2 Z_{\xi\eta}}{\partial \mathbf{u}_p \partial \mathbf{u}_q} = \begin{pmatrix} \left\|\frac{\partial^2 Z_{\xi\eta}}{\partial E_{\alpha p}\partial E_{\beta q}}\right\| & \left\|\frac{\partial^2 Z_{\xi\eta}}{\partial E_{\alpha p}\partial H_{\beta q}}\right\| \\ \left\|\frac{\partial^2 Z_{\xi\eta}}{\partial H_{\alpha p}\partial E_{\beta q}}\right\| & \left\|\frac{\partial^2 Z_{\xi\eta}}{\partial H_{\alpha p}\partial H_{\beta q}}\right\| \end{pmatrix} = \begin{pmatrix} \|0\| & \left\|-\delta_{\xi\alpha} Q_{p\beta} Q_{q\eta}\right\| \\ \left\|-\delta_{\xi\beta} Q_{q\alpha} Q_{p\eta}\right\| & \left\|Z_{\xi\beta} Q_{q\alpha} Q_{p\eta} + Z_{\xi\alpha} Q_{p\beta} Q_{q\eta}\right\| \end{pmatrix}, \tag{82}$$

where $\|\cdot\|$ stands for a 2×2 matrix, $\xi, \eta =$‘x’, ‘y’, $\alpha, \beta, p, q = 1, 2$. Thus, the adjoint field $\mathbf{G}^{e\cdot}(\mathbf{J}^\Psi)$ from eq. (66) is

$$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^\Psi(\boldsymbol{\nu}'\mathbf{1}_{V_k}, \omega)\right) = \sum_{\substack{a\in\mathcal{S}\\ \xi,\eta=1,2}} \left.\frac{\left(Z_{\xi\eta}^{\mathrm{pred}} - Z_{\xi\eta}^{\mathrm{exp}}\right)^*}{|\Delta Z_{\xi\eta}^{\mathrm{exp}}|^2}\right|_{\mathbf{s}_a,\omega} \mathbf{G}^{e\cdot}\left.\left(\frac{\partial^2 Z_{\xi\eta}}{\partial \mathbf{u}_p \partial \mathbf{u}_q}\cdot \mathbf{G}^{\cdot e}\left(\boldsymbol{\nu}'\mathbf{1}_{V_k}\,\mathbf{E}_q\right)\delta_{\mathbf{s}_a}\right)\right|_\omega. \tag{83}$$

Finally, the Hessian for MT data misfit can be found from eq. (61), (62) and (63) where $\mathbf{G}^{e\cdot}(\mathbf{J}_p^M)$, $\mathbf{G}^{e\cdot}(\mathbf{J}_p^B)$ and $\mathbf{G}^{e\cdot}(\mathbf{J}_p^\Psi)$ are from eqs (80), (81) and (83).

Table 5 provides a number of forward modellings needed to calculate Hessian-vector product for K times. As it is seen from the table, single Hessian-vector product can be calculated for a price of $O(N_\Omega N_{\mathcal{S}})$ forward problem runs. Moreover, if such a product is calculated multiple times the price drops down to $4N_\Omega$ per product. And again we notice that computation of Hessian is nothing but calculation of Hessian-vector product for $K = N_{\mathcal{M}}$ times with respective vectors $\mathbf{a}_k = \mathbf{1}_{V_k}$, $k = 1, \ldots, N_{\mathcal{M}}$.

6 CONCLUSIONS

We present a general formalism for the efficient calculation of the second derivatives of EM frequency-domain responses and the second derivatives of the misfit (Hessian matrix) with respect to variations of 3-D conductivity. We consider this paper as a prelude to the development of quantitative resolution schemes in EM problems.

We also show how this technique can be implemented to calculate multiple Hessian-vector products very efficiently. This opens an avenue to implement truncated Newton optimization methods for 3-D EM inversions, and more important, a way to implement a low-rank approximation of the Hessian matrix in the line as it was discussed in section 2.4 in the paper by Martin *et al.* (2012). As the further steps of the algorithm a stochastic approach can be applied to 3-D EM inversion in order to study the domain of all models that have approximately the same misfit (equivalence domain), or a local deterministic approach can be applied in order to quantify the uncertainty of a model that has been acquired by some inversion technique.

The formalism uses the adjoint sources approach and allows one to work with responses that arise in EM problem set-ups either with natural- or controlled-source excitations. The formalism allows for various types of parametrization of the 3-D conductivity distribution. Using this methodology one can readily obtain appropriate formulae for the specific sounding methods. To illustrate the concept we provide such formulae for two EM techniques: magnetotellurics and controlled-source sounding with vertical magnetic dipole as a source.

Applying of the developed formalism to practical scenarios is intentionally beyond the scope of the paper but will be the subject of a subsequent study.

ACKNOWLEDGEMENTS

We wish to thank William Lowrie who helped us to improve the English presentation of this paper. We would like to thank the Editors and two anonymous reviewers for their useful comments and suggestions that have greatly improved the manuscript. This work has been supported by the European Space Agency through ESTEC contract No. 4000102140/10/NL/JA and in part by the Russian Foundation for Basic Research under grant No. 13-05-12111. Oleg Pankratov acknowledges the support of ETH during his stay as a visiting Professor at ETH Zurich.

REFERENCES

Avdeev, D. & Avdeeva, A., 2009. 3D magnetotelluric inversion using a limited-memory quasi-Newton optimization, *Geophysics,* **74**(3), F45–F57.

Egbert, G.D. & Kelbert, A., 2012. Computational recipes for electromagnetic inverse problems, *Geophys. J. Int.,* **189**(1), 251–267.

Epanomeritakis, I., Akcelik, V., Ghattas, O. & Bielak, J., 2008. A Newton-CG method for large-scale three-dimensional elastic full waveform seismic inversion, *Inverse Problems,* **24,** 1–26.

Fichtner, A. & Trampert, J., 2011. Hessian kernels of seismic data functionals based upon adjoint techniques, *Geophys. J. Int.,* **185,** 775–798.

Haber, E., Ascher, U. & Oldenburg, D., 2004. Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach, *Geophysics,* **69**(5), 1216–1228.

Kelbert, A., Egbert, G. & Schultz, A., 2008. A nonlinear conjugate 3-D inversion of global induction data. Resolution studies, *Geophys. J. Int.,* **173,** 365–381.

Koyama, T., 2001. A study on the electrical conductivity of the mantle by voltage measurements of submarine cables, *PhD thesis,* University of Tokyo.

Kuvshinov, A. & Semenov, A., 2012. Global 3-D imaging of mantle electrical conductivity based on inversion of observatory C-responses—I. An approach and its verification, *Geophys. J. Int.,* **189,** 1335–1352.

Mackie, R. & Madden, T., 1993. Three-dimensional magnetotelluric inversion using conjugate gradients, *Geophys. J. Int.,* **115**(1), 215–229.

Martin, J., Wilcox, L.C., Burstedde, C. & Ghattas, O., 2012. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion, *SIAM J. Sci. Comput.,* **34**(3), A1460–A1487.

Metivier, L., Brossier, R., Virieux, J. & Operto, S., 2013. Full waveform inversion and the truncated Newton method, *SIAM J. Sci. Comput.,* **35,** 401–437.

Nash, S., 2000. A survey of truncated-Newton methods, *J. Comput. Appl. Math.,* **35,** 45–59.

Newman, G. & Alumbaugh, D., 2000. Three-dimensional magnetotelluric inversion using non-linear conjugate gradients, *Geophys. J. Int.,* **140**(2), 410–424.

Newman, G.A. & Hoversten, G.M., 2000. Solution strategies for two- and three-dimensional electromagnetic inverse problems, *Inverse Problems,* **16,** 1357–1375.

Pankratov, O. & Kuvshinov, A., 2010. General formalism for the efficient calculation of derivatives of EM frequency-domain responses and derivatives of the misfit, *Geophys. J. Int.,* **181**(1), 229–249.

Santosa, F. & Symes, W.W., 1988. Computation of the Hessian for least-squares solutions of inverse problems of reflection seismology, *Inverse Problems,* **4,** 211–213.

Sasaki, Y. & Meju, M., 2006. Three-dimensional joint inversion for magnetotelluric resistivity and static shift distributions in complex media, *J. geophys. Res.,* **111**(B5), doi:10.1029/2005JB004009.

Siripunvaraporn, W., Egbert, G., Lenbury, Y. & Uyeshima, M., 2005. Three-dimensional magnetotelluric inversion: data-space method, *Phys. Earth planet. Int.,* **150**(1–3), 3–14.

Zhang, L., Koyama, T., Utada, H., Yu, P. & Wang, J., 2012. A regularized three-dimensional magnetotelluric inversion with a minimum gradient support constraint, *Geophys. J. Int.,* **189**(1), 296–316.

APPENDIX A: MULTILINEAR CONTRACTIONS

In this appendix, we define contraction operations used in the main body of the paper. Let us consider a set of functions $u_k(\boldsymbol{\sigma})$, $k = 1, \ldots, K$, where $\boldsymbol{\sigma}$ is determined by eq. (17). We also introduce $\mathrm{d}\boldsymbol{\sigma}$ as

$$\mathrm{d}\boldsymbol{\sigma} = (\mathrm{d}\sigma_1, \ \ldots, \ \mathrm{d}\sigma_{N_{\mathcal{M}}})^T. \tag{A1}$$

The first differential of u_k with respect to $\boldsymbol{\sigma}$ is written as

$$\mathrm{d}u_k(\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\sigma}) = \sum_{m=1}^{N_{\mathcal{M}}} \frac{\partial u_k}{\partial \sigma_m} \cdot \mathrm{d}\sigma_m = \left(\frac{\partial u_k}{\partial \sigma_1}, \ldots, \frac{\partial u_k}{\partial \sigma_{N_{\mathcal{M}}}} \right) \begin{pmatrix} \mathrm{d}\sigma_1 \\ \vdots \\ \mathrm{d}\sigma_{N_{\mathcal{M}}} \end{pmatrix} = \left\langle \frac{\partial u_k}{\partial \boldsymbol{\sigma}} \,\middle|\, \mathrm{d}\boldsymbol{\sigma} \right\rangle, \tag{A2}$$

where we denote $(\frac{\partial u_k}{\partial \sigma_1}, \ldots, \frac{\partial u_k}{\partial \sigma_{N_{\mathcal{M}}}})^T$ as $\frac{\partial u_k}{\partial \boldsymbol{\sigma}}$. We call RHS of the latter equation a 'bilinear scalar-valued contraction'. By introducing $\mathbf{u}$ as

$$\mathbf{u} = \{u_k\}_{k=1,\ldots,K} \ , \tag{A3}$$

we can write vector differential $\mathrm{d}\mathbf{u}$ as

$$\mathrm{d}\mathbf{u}(\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\sigma}) = \{\mathrm{d}u_k(\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\sigma})\}_{k=1,\ldots,K} = \left\langle \frac{\partial \mathbf{u}}{\partial \boldsymbol{\sigma}} \,\middle|\, \mathrm{d}\boldsymbol{\sigma} \right\rangle, \tag{A4}$$

where $\frac{\partial \mathbf{u}}{\partial \boldsymbol{\sigma}}$ is a matrix with elements $\frac{\partial u_k}{\partial \sigma_m}$, $k = 1, \ldots, K$, $m = 1, \ldots, N_{\mathcal{M}}$. We call RHS of the latter equation a 'bilinear vector-valued contraction'.

Let us now calculate the differential of a composite function $F[\mathbf{u}(\boldsymbol{\sigma})]$

$$\mathrm{d}F[\mathbf{u}(\boldsymbol{\sigma})] = \sum_{k=1}^{K} \frac{\partial F}{\partial u_k} \cdot \mathrm{d}u_k = \left\langle \frac{\partial F}{\partial \mathbf{u}} \,\middle|\, \mathrm{d}\mathbf{u} \right\rangle. \tag{A5}$$

Substituting eq. (A4) in eq. (A5) we get the differential for the composite function

$$\mathrm{d}F(\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\sigma}) = \sum_{k=1}^{K} \sum_{m=1}^{N_{\mathcal{M}}} \frac{\partial F}{\partial u_k} \cdot \frac{\partial u_k}{\partial \sigma_m} \cdot \mathrm{d}\sigma_m = \left\langle \frac{\partial F}{\partial \mathbf{u}} \,\middle|\, \frac{\partial \mathbf{u}}{\partial \boldsymbol{\sigma}} \,\middle|\, \mathrm{d}\boldsymbol{\sigma} \right\rangle. \tag{A6}$$

We call RHS of the latter equation a 'trilinear scalar-valued contraction'. Note that for this contraction the following associativity rule holds

$$\left\langle \left\langle \frac{\partial F}{\partial \mathbf{u}} \,\middle|\, \frac{\partial \mathbf{u}}{\partial \boldsymbol{\sigma}} \right\rangle \,\middle|\, \mathrm{d}\boldsymbol{\sigma} \right\rangle = \left\langle \frac{\partial F}{\partial \mathbf{u}} \,\middle|\, \left\langle \frac{\partial \mathbf{u}}{\partial \boldsymbol{\sigma}} \,\middle|\, \mathrm{d}\boldsymbol{\sigma} \right\rangle \right\rangle = \left\langle \frac{\partial F}{\partial \mathbf{u}} \,\middle|\, \frac{\partial \mathbf{u}}{\partial \boldsymbol{\sigma}} \,\middle|\, \mathrm{d}\boldsymbol{\sigma} \right\rangle. \tag{A7}$$

Another example of trilinear scalar-valued contraction is the second differential $d^2 u_k(\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\eta})$ which can be presented as

$$d^2 u_k(\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\eta}) = \left(\mathrm{d}\sigma_1, \ \ldots, \ \mathrm{d}\sigma_{N_{\mathcal{M}}}\right) \begin{pmatrix} \partial^2 u_k/\partial\sigma_1\partial\sigma_1 & \ldots & \partial^2 u_k/\partial\sigma_1\partial\sigma_{N_{\mathcal{M}}} \\ \vdots & \ldots & \vdots \\ \partial^2 u_k/\partial\sigma_{N_{\mathcal{M}}}\partial\sigma_1 & \ldots & \partial^2 u_k/\partial\sigma_{N_{\mathcal{M}}}\partial\sigma_{N_{\mathcal{M}}} \end{pmatrix} \begin{pmatrix} \mathrm{d}\eta_1 \\ \vdots \\ \mathrm{d}\eta_{N_{\mathcal{M}}} \end{pmatrix} = \left\langle \mathrm{d}\boldsymbol{\sigma} \,\middle|\, \frac{\partial^2 u_k}{\partial \boldsymbol{\sigma}^2} \,\middle|\, \mathrm{d}\boldsymbol{\eta} \right\rangle. \tag{A8}$$

Finally, vector second differential is a set of its scalar components

$$d^2 \mathbf{u}(\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\eta}) = \left\{ d^2 u_k(\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\eta}) \right\}_{k=1,\ldots,K} = \left\langle \mathrm{d}\boldsymbol{\sigma} \,\middle|\, \frac{\partial^2 \mathbf{u}}{\partial \boldsymbol{\sigma}^2} \,\middle|\, \mathrm{d}\boldsymbol{\eta} \right\rangle. \tag{A9}$$

We call the RHS in the latter equation a 'trilinear vector-valued contraction'.

Note that all the above contractions are presented for the finite-dimensional case (17). One can readily obtain the corresponding formulas (via integrals) for continuous case, $\boldsymbol{\sigma} = \sigma(\mathbf{r})$. For example, eq. (A2) in the continuous case reads

$$\delta u^k(\boldsymbol{\sigma}, \mathrm{d}\boldsymbol{\sigma}) = \left\langle \frac{\partial u_k}{\partial \boldsymbol{\sigma}} \,\middle|\, \mathrm{d}\boldsymbol{\sigma} \right\rangle = \int_{\mathbb{R}^3} \frac{\partial u^k}{\partial \sigma(\mathbf{r})} \cdot \delta\sigma(\mathbf{r})\, \mathrm{d}v(\mathbf{r}). \tag{A10}$$

APPENDIX B: CALCULATING DERIVATIVES OF THE MAXWELL's OPERATOR

B1 Calculating the term $\frac{\partial \mathbf{L}(\mathbf{u},\sigma)}{\partial \sigma}$

In this appendix, we calculate $\frac{\partial \mathbf{L}(\mathbf{u},\sigma)}{\partial \sigma}$. It should be a tensor that is able to be contracted with an arbitrary variation of $\boldsymbol{\sigma}$. Using the following equation:

$$\frac{\partial}{\partial \boldsymbol{\sigma}} \begin{pmatrix} -\boldsymbol{\sigma} & 0 \\ 0 & 0 \end{pmatrix} = -\begin{pmatrix} \mathbb{1} & 0 \\ 0 & 0 \end{pmatrix} \delta\left(\mathbf{r} - \mathbf{r}'\right), \tag{B1}$$

from eq. (11) we obtain

$$\frac{\partial \mathbf{L}}{\partial \boldsymbol{\sigma}} = \frac{\partial \mathbb{L}}{\partial \boldsymbol{\sigma}} \mathbf{u} = \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{\sigma}} \begin{pmatrix} -\mathbb{1} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{u} = -\delta\left(\mathbf{r} - \mathbf{r}'\right) \begin{pmatrix} \mathbf{E} \\ 0 \end{pmatrix}. \tag{B2}$$

Here we used a projection of EM field, $\mathbf{u} = (\mathbf{E}, \mathbf{H})$, onto its electric component, $\mathbf{E}$. Contracting (B2) with an arbitrary conductivity variation $\delta\sigma(\mathbf{r})$, we have

$$\left\langle \frac{\partial \mathbf{L}}{\partial \boldsymbol{\sigma}} \,\middle|\, \delta\boldsymbol{\sigma} \right\rangle = \begin{pmatrix} -\delta\sigma(\mathbf{r})\, \mathbf{E}(\mathbf{r}) \\ 0 \end{pmatrix}. \tag{B3}$$

B2 Calculating the term $\frac{\partial^2 \mathbf{L}(\mathbf{u},\sigma)}{\partial \mathbf{u} \partial \sigma}$

In this appendix we calculate $\frac{\partial^2 \mathbf{L}(\mathbf{u},\sigma)}{\partial \mathbf{u} \partial \sigma}$. It should be a tensor that can be contracted with an arbitrary variation of $\mathbf{u}$ and an arbitrary variation of $\boldsymbol{\sigma}$

$$\frac{\partial^2 \mathbf{L}}{\partial \mathbf{u} \partial \boldsymbol{\sigma}} = \frac{\partial^2}{\partial \mathbf{u} \partial \boldsymbol{\sigma}} \mathbb{L}\mathbf{u} = \frac{\partial \mathbb{L}}{\partial \boldsymbol{\sigma}} \frac{\partial \mathbf{u}}{\partial \mathbf{u}} = \frac{\partial}{\partial \boldsymbol{\sigma}} \begin{pmatrix} -\boldsymbol{\sigma} & \nabla\times \\ \nabla\times & -i\omega\mu \end{pmatrix} \frac{\partial \mathbf{u}}{\partial \mathbf{u}}. \tag{B4}$$

The first cofactor of the latter expression is found from eq. (B1). The 2×2-matrix in the RHS denotes a tensor that can be multiplied by the 6-D EM field $(\mathbf{E}, \mathbf{H})^T$ and produces the result $(\mathbf{E}, 0)^T$. The second co-factor in the RHS of eq. (B4) is

$$\frac{\partial \mathbf{u}}{\partial \mathbf{u}} = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix} \delta\left(\mathbf{r} - \mathbf{r}''\right). \tag{B5}$$

The 2×2 matrix in the RHS denotes an identical operator in the space of 6-D EM values $\mathbb{C}^3 \times \mathbb{C}^3$. Combining eqs (B4), (B1) and (B5) we arrive at

$$\frac{\partial^2 \mathbf{L}}{\partial \mathbf{u} \partial \boldsymbol{\sigma}} = -\delta\left(\mathbf{r} - \mathbf{r}'\right) \begin{pmatrix} \mathbb{1} & 0 \\ 0 & 0 \end{pmatrix} \delta\left(\mathbf{r} - \mathbf{r}''\right). \tag{B6}$$

By contracting an arbitrary model variation $\delta \mathbf{m}(\mathbf{r})$ and an arbitrary EM field variation $\delta \mathbf{u}(\mathbf{r})$ with the latter equation we have

$$\left\langle \delta \mathbf{m} \left| \frac{\partial^2 \mathbf{L}}{\partial \boldsymbol{\sigma} \partial \mathbf{u}} \right| \delta \mathbf{u} \right\rangle = \begin{pmatrix} -\delta m(\mathbf{r})\, \delta \mathbf{E}(\mathbf{r}) \\ 0 \end{pmatrix}. \tag{B7}$$

APPENDIX C: THE FIRST DIFFERENTIAL FORMULA FOR A RESPONSE FUNCTION

Here we prove formula (43). We recall that θ_g can be written as

$$\theta_g(\mathbf{m}) = \Psi_g\left(\mathbf{u}_1(\mathbf{m}, \mathbf{r}_g, \omega_g), \mathbf{u}_2(\mathbf{m}, \mathbf{r}_g, \omega_g), \ldots, \mathbf{u}_{N_\mathcal{P}}(\mathbf{m}, \mathbf{r}_g, \omega_g)\right), \tag{C1}$$

where $\mathbf{u}_p$ is an EM field due to the pth polarization source from eq. (13). Assuming that response function Ψ_g is a complex-differentiable function of all its arguments $\mathbf{u}_p$, $p \in \mathcal{P}$, and differentiating eq. (C1), we have

$$\delta\theta_g(\mathbf{m}) = \sum_{p \in \mathcal{P}} \frac{\partial \Psi_g}{\partial \mathbf{u}_p} \cdot \delta \mathbf{u}_p \bigg|_{\mathbf{r}_g, \omega_g} = \sum_{p \in \mathcal{P}} \left\langle \frac{\partial \Psi_g}{\partial \mathbf{u}_p} \delta_{\mathbf{r}_g} \middle| \delta \mathbf{u}_p \right\rangle \bigg|_{\omega_g}. \tag{C2}$$

Here we converted substitution of a fixed argument to the form an integral with the Dirac's delta function, $F(\mathbf{r}_g) = \int_{\mathbb{R}^3} F(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}_g)\,\mathrm{d}v$. Next, we invoke formula (39) as well as the reciprocity property of Green's function [*cf.* eqs (A1)–(A2) in Pankratov & Kuvshinov (2010)]

$$\langle\, \mathbf{G}^{e\cdot}(\mathbf{a}) \,|\, \mathbf{b} \,\rangle = \langle\, \mathbf{a} \,|\, \mathbf{G}^{\cdot e}(\mathbf{b}) \,\rangle \tag{C3}$$

and get

$$\sum_{p \in \mathcal{P}} \left\langle \frac{\partial \Psi_g}{\partial \mathbf{u}_p} \delta_{\mathbf{r}_g} \middle| \delta \mathbf{u}_p \right\rangle \bigg|_{\omega_g} = \sum_{p \in \mathcal{P}} \left\langle \frac{\partial \Psi_g}{\partial \mathbf{u}_p} \delta_{\mathbf{r}_g} \middle| \mathbf{G}^{\cdot e}\left(\delta\boldsymbol{\sigma}\, \mathbf{E}_p\right) \right\rangle \bigg|_{\omega_g} = \sum_{p \in \mathcal{P}} \left\langle \delta\boldsymbol{\sigma}\, \mathbf{E}_p \middle| \mathbf{G}^{e\cdot}\left(\frac{\partial \Psi_g}{\partial \mathbf{u}_p} \delta_{\mathbf{r}_g}\right) \right\rangle \bigg|_{\omega_g}. \tag{C4}$$

Deciphering the bilinear form $\langle\, \cdot \mid \cdot\, \rangle$ we finally obtain

$$\delta\theta_g(\mathbf{m}) = \sum_{p \in \mathcal{P}} \int_{\mathbb{R}^3} \delta\sigma(\mathbf{r})\, \mathbf{E}_p(\mathbf{r}) \cdot \mathbf{G}^{e\cdot}\left(\frac{\partial \Psi_g}{\partial \mathbf{u}_p} \delta_{\mathbf{r}_g}\right)\bigg|_{\mathbf{r}} \mathrm{d}v(\mathbf{r}) \bigg|_{\omega_g}, \tag{C5}$$

where the dot operation involves the $\mathbb{C}$-bilinear product $\mathbf{E}(\mathbf{r}) \cdot \mathbf{g}(\mathbf{r}) = E_x(\mathbf{r})g_x(\mathbf{r}) + E_y(\mathbf{r})g_y(\mathbf{r}) + E_z(\mathbf{r})g_z(\mathbf{r})$.

APPENDIX D: EXTREMAL BOUNDS IN THE MODEL SPACE

Let point $\mathbf{m}_0$ be a stationary point of the penalty function $\beta(\mathbf{m})$

$$\nabla\beta(\mathbf{m}_0) = 0. \tag{D1}$$

Let $\Delta\beta$ be a perturbation of $\beta(\mathbf{m}_0)$. We study the allowable extremal bounds in the model space that are permitted by the value of $\Delta\beta$

$$\mathbf{m}\ : \beta(\mathbf{m}) \leq \beta(\mathbf{m}_0) + \Delta\beta. \tag{D2}$$

Let us define function f as follows:

$$f(\Delta\mathbf{m}) = \beta(\mathbf{m}_0 + \Delta\mathbf{m}) - \beta(\mathbf{m}_0), \tag{D3}$$

and introduce a quadratic approximation of function f as

$$g(\Delta\mathbf{m}) = f(0) + \nabla f(0)^T \Delta\mathbf{m} + \frac{1}{2}\Delta\mathbf{m}^T \mathbf{H}_0 \Delta\mathbf{m} = \frac{1}{2}\Delta\mathbf{m}^T \mathbf{H}_0 \Delta\mathbf{m}, \tag{D4}$$

where $f(0) = 0$ due to eq. (D3), and $\nabla f(0) = 0$ due to eq. (D1). Here for brevity we denote $\mathbf{H}_0 = \mathrm{Hess}(\mathbf{m}_0)$.

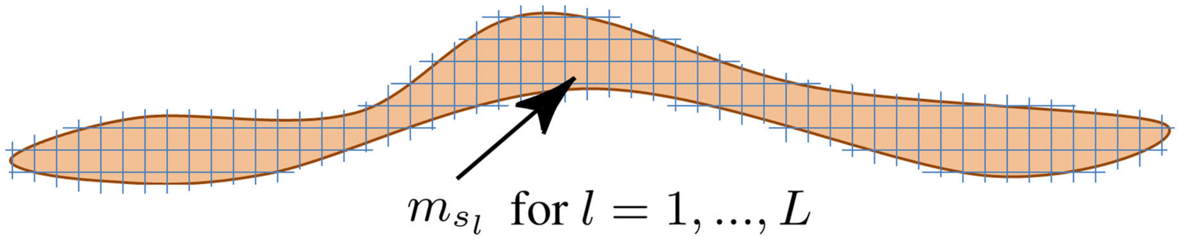

Figure D1. A spatial domain V_d and a set of variables $m_{s_1}, \ldots, m_{s_L}$ that determine the conductivity distribution within domain V_d. We choose the domain V_d as a volume where we expect certain behaviour of conductivity (e.g. an oil trap). We may expect this behaviour from either some *a priori* data or from some previous inversion results as well, and we wish to estimate the limits of the average conductivity (or the parameter m) in the volume.

Let us choose a subdomain $V_d \subseteq V^{\mathrm{inv}}$ (see Fig. D1) of the inversion domain $V^{\mathrm{inv}} = \bigcup_{k=1}^{N_{\mathcal{M}}} V_k$ (see eq. 16). The subdomain V_d is an object that we are interested in studying with respect to variability of the average (over V_d) of the conductivity (or, more generally, the average value of the parameter m) given the responses fit the observed data within the perturbation $\Delta\beta$. Let us define vector $\mathbf{q}$ as follows:

$$\mathbf{q} = (q_1, \ldots, q_{N_{\mathcal{M}}})^T, \quad q_k = \mathrm{Volume}(V_d \cap V_k), \quad \text{for } k = 1, \ldots, N_{\mathcal{M}}. \tag{D5}$$

Then the average value of parameter m over V_d is

$$\mathbf{q}^T\mathbf{m} = \sum_{k=1}^{N_{\mathcal{M}}} q_k m_k. \tag{D6}$$

Next, we want to maximize the latter expression, given that the misfit satisfies eq. (D2). Solving this conditional extremum problem with the Lagrangian method

$$\begin{cases} \nabla g || \mathbf{q} \\ g(\Delta\mathbf{m}) = \Delta\beta, \end{cases} \tag{D7}$$

we obtain

$$\begin{cases} \mathbf{H}_0\Delta\mathbf{m} = \lambda\mathbf{q} \\ \Delta\mathbf{m}^T\mathbf{H}_0\Delta\mathbf{m} = 2\Delta\beta \end{cases} \Rightarrow \begin{cases} \Delta\mathbf{m} = \lambda\mathbf{H}_0^{-1}\mathbf{q} \\ \lambda^2\,\mathbf{q}^T\mathbf{H}_0^{-1}\mathbf{q} = 2\Delta\beta \end{cases}. \tag{D8}$$

From the latter equation we determine λ as

$$\lambda = \pm\sqrt{\frac{2\Delta\beta}{\mathbf{q}^T\mathbf{H}_0^{-1}\mathbf{q}}}. \tag{D9}$$

Finally, for $\Delta\mathbf{m}$ for which the desired inequality

$$\beta\,(\mathbf{m}_0 \pm \Delta\mathbf{m}) \le \beta\,(\mathbf{m}_0) + \Delta\beta, \tag{D10}$$

holds, we have the extremal bounds as follows:

$$\Delta\mathbf{m} = \mathbf{H}_0^{-1}\mathbf{q}\sqrt{\frac{2\Delta\beta}{\mathbf{q}^T\mathbf{H}_0^{-1}\mathbf{q}}}. \tag{D11}$$

Here we assumed that matrix $\mathbf{H}_0$ is positively definite. In the case when the Hessian matrix is diagonal and the gridding is coarse enough so that the volume V_d is represented by a single mesh cell V_k, then the formula for the perturbation $\Delta\mathbf{m}_k$ reads

$$\Delta\mathbf{m}_k = \mathbf{e}_k\sqrt{\frac{2\Delta\beta}{\mathrm{Hess}(\mathbf{m}_0)_{k,k}}}, \quad k = 1, 2, \ldots N_{\mathcal{M}}, \tag{D12}$$

where $\mathbf{e}_k$ is a unit vector of dimension $N_{\mathcal{M}}$ with a non-zero entry at kth position, and where $\Delta\mathbf{m}_k$ is such a model perturbation that delivers a maximum deviation in the kth component provided that quadratic approximation (D4) of the misfit is within the perturbation $\Delta\beta$ (see Fig. D2). In eq. (D12) we redenote $\mathrm{Hess}(\mathbf{m}_0) = \mathbf{H}_0$. If the Hessian matrix is not diagonal a more sophisticated formula holds

$$\Delta\mathbf{m}_k = \mathrm{Hess}(\mathbf{m}_0)^{-1}\mathbf{e}_k\sqrt{\frac{2\Delta\beta}{\left(\mathrm{Hess}(\mathbf{m}_0)^{-1}\right)_{k,k}}}. \tag{D13}$$

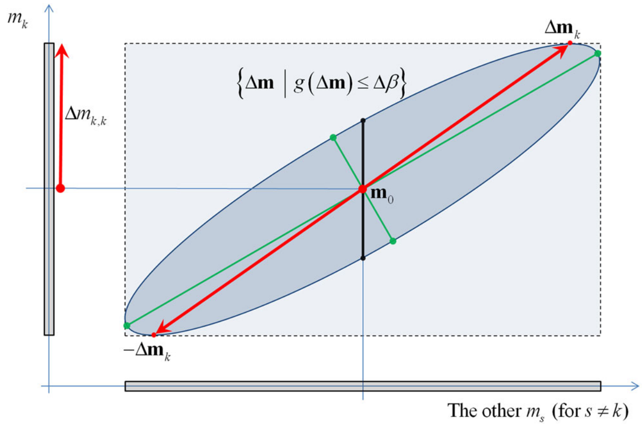

Figure D2. A schematic illustration of eq. (D13). Big red point is the model $\mathbf{m}_0$ that is a stationary point of the penalty function. Vectors $\pm\Delta\mathbf{m}_k$ (as shown in red ink) deliver a maximum and a minimum to the parameter m_k within the ellipsoid $\Delta\mathbf{m}$: $g(\Delta\mathbf{m}) \leq \Delta\beta$. The value of $\Delta m_{k,\,k}$ delivers a majorant to the uncertainty in m_k. One should not confuse the vectors $\pm\Delta\mathbf{m}_k$ with the vectors (as shown in black ink) that deliver maximum and minimum to the parameter m_k by only varying m_k. The latter deliver a minorant to the uncertainty in m_k. Green orthogonal axes are the eigenvectors of the Hessian of $g(\Delta\mathbf{m})$. The lengths of the axes are determined by the eigenvalues of the Hessian and could deliver the most precise estimation to the model uncertainty.

APPENDIX E: HESSIAN OF THE MISFIT FOR COMPLEX-VALUED CONDUCTIVITIES

Using the results of Section 4.4 one can readily generalize the concept and get the misfit Hessian matrix for complex-valued conductivity, $\sigma(\mathbf{r}, \omega)$. In this case the entries of Hessian matrix look as follows:

$$\begin{pmatrix} \frac{\partial^2 \beta_d}{\partial \mathrm{Re}\, m_k \partial \mathrm{Re}\, m_l} & \frac{\partial^2 \beta_d}{\partial \mathrm{Re}\, m_k \partial \mathrm{Im}\, m_l} \\ \frac{\partial^2 \beta_d}{\partial \mathrm{Im}\, m_k \partial \mathrm{Re}\, m_l} & \frac{\partial^2 \beta_d}{\partial \mathrm{Im}\, m_k \partial \mathrm{Im}\, m_l} \end{pmatrix} = \begin{pmatrix} \mathrm{Re}(\mathcal{H}^A_{kl} + \mathcal{H}^L_{kl}) & \mathrm{Im}(\mathcal{H}^A_{kl} - \mathcal{H}^L_{kl}) \\ -\,\mathrm{Im}(\mathcal{H}^A_{kl} + \mathcal{H}^L_{kl}) & \mathrm{Re}(\mathcal{H}^A_{kl} - \mathcal{H}^L_{kl}) \end{pmatrix}, \tag{E1}$$

where $\mathcal{H}^A_{kl}$ and $\mathcal{H}^L_{kl}$ are determined in eq. (62).